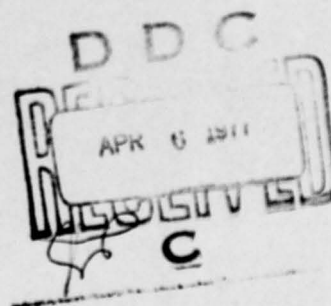RADC-TR-76-380
In-House Technical Report
December 1976

AGRICULTURAL CROP YIELD PREDICTION UTILIZING NARROWBAND
MULTISPECTRAL, TEMPORALLY-REGISTERED IMAGERY
(A FEASIBILITY STUDY'S TECHNICAL REPORT AND RECOMMENDATIONS)

1/Lt Gregory B. Pavlin

DDC

APR 6 1977

C

Approved for public release;
distribution unlimited.

ROME AIR DEVELOPMENT CENTER
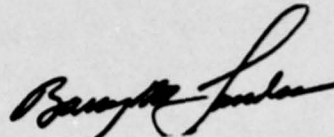**AIR FORCE SYSTEMS COMMAND**
GRIFFISS AIR FORCE BASE, NEW YORK 13441

This study was accomplished using only extra duty manhours. Over a 3-year period, the investigator, 1/Lt Gregory Pavlin, voluntarily expended approximately 2000 manhours in the process of completing the investigation.

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public including foreign nations.
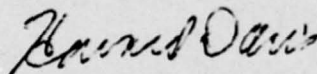
This report has been reviewed and is approved for publication.

APPROVED:
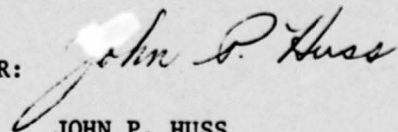
BARRY M. LANDSON, Major, USAF
Chief, Image Systems Branch

APPROVED:

HOWARD DAVIS
Technical Director
Intelligence & Reconnaissance Division

FOR THE COMMANDER:

JOHN P. HUSS
Acting Chief, Plans Office

AGRICULTURAL CROP YIELD PREDICTION

UTILIZING

NARROWBAND MULTISPECTRAL-TEMPORALLY REGISTERED IMAGERY

A FEASIBILITY STUDY'S REPORT, CONCLUSIONS AND RECOMMENDATIONS

ROME AIR DEVELOPMENT CENTER

AIR FORCE SYSTEMS COMMAND

RADC/IRRE (LT. GREGORY B. PAVLIN)

GRIFFISS A.F.B., ROME, NEW YORK 13441

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC-TR-76-380	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AGRICULTURAL CROP YIELD PREDICTION UTILIZING NARROWBAND MULTISPECTRAL, TEMPORALLY-REGISTERED IMAGERY (A FEASIBILITY STUDY'S TECHNICAL REPORT AND RECOMMENDATIONS).		5. TYPE OF REPORT & PERIOD COVERED In-House Report June 1974 – June 1976
		6. PERFORMING ORG. REPORT NUMBER N/A
7. AUTHOR(s) 1/Lt Gregory B. Pavlin		8. CONTRACT OR GRANT NUMBER(s) N/A
9. PERFORMING ORGANIZATION NAME AND ADDRESS Rome Air Development Center (IRRE) Griffiss Air Force Base, New York 13441		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62702F 62441051
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRRE) Griffiss Air Force Base, New York 13441		12. REPORT DATE December 1976
		13. NUMBER OF PAGES 405
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Rome Air Development Center (IRRE) Griffiss Air Force Base, New York 13441		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

This study was accomplished using only extra duty manhours. Over a 3-year period, the investigator, 1/Lt Gregory Pavlin, voluntarily expended approximately 2000 manhours in the process of completing the investigation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Agriculture	Stress Detection	Agricultural Crops
Crop Yield	Crop Identification	Regression Analysis
Remote Sensing	Multispectral Photography	Multivariate Analysis
Image Processing	Pattern Recognition	Yield
Digital Image Processing	Radiometry	(see reverse)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Selected narrowband multispectral photography, dedicated to detecting levels of stress within corn, was collected regularly at low and high altitudes over eight large agricultural test sites in Central Pennsylvania during the period of April thru October of 1974. A spectroradiometry survey was conducted at the test sites during three intervals of the 1974 corn growth cycle to verify the selection of the multispectral camera system's filters. Superlative ground and image truth were collected regularly to supplement the multispectral

over

APR 6 1977

309050

photography of the test sites. After analyzing the multispectral photography with additive color analysis techniques, selected multispectral photography of three test sites, spanning the entire growth cycle, was digitized and registered, both spectrally and temporally, using the image processing resources of the Rome Air Development Center's Image Processing Facility. By analyzing the digital, temporally-registered, multispectral image data with weather and ground truth data, utilizing the techniques of multivariate and regression analysis, the feasibility of agricultural crop yield prediction was investigated.

The feasibility of timely (as early as pollination phase) and accurate crop yield prediction (87% of the total crop yield data's variance was accounted for with a 90% accuracy) of fourteen small corn fields, located within the boundaries of the three multispectrally-imaged test sites, was demonstrated utilizing a yield prediction equation that was solely dependent upon digital, temporally-registered, multispectral image data. Furthermore, relative to the image data, the weather data was observed to be an insignificant predictor of crop yield.

Subsidiary image processing studies of this program investigated the feasibility of semiautomatic crop identification. The multispectral imagery demonstrated itself to be excellent data for semiautomatically generating digital maps of stress within multispectrally-imaged fields of corn. The average yield of the three agricultural test sites could be accurately measured using digital stress maps produced from late season imagery and crop yield sampling data collected during the ground truth program. Only marginal semiautomatic corn identification accuracies were achieved using image data dedicated to stress level discrimination within corn.

The overall program was significant because the findings indicate that timely and accurate crop yield prediction may be possible (discounting catastrophic events; i.e.,: floods, fires, etc.) using temporally-registered, multispectral imagery. However, though the findings are promising, the relationships will have to be verified through future studies. If the technology is proven, the economic intelligence community would have an ideal method for determining the agricultural status of foreign countries. Timely and reliable crop yield information would diminish the risks of foreign grain deals; would add stability and predictability to grain price indices; and could conceivably forewarn grain producers to take corrective action so that their losses would be minimized when poor crops were predicted.

Block 19:   Narrowband Filters
             Temporal Registration

FOREWARD

During 1973, personnel of the Rome Air Development
Center (RADC) demonstrated the feasibility of accurate
remote crop identification utilizing a narrowband multi-
spectral camera and associated additive color display
technology (Haynes, 23). This significant accomplishment
encouraged RADC personnel to further study this unique
technology's potential for solving the military problem
of detecting camouflaged tactical targets. Consequently,
during 1974, RADC personnel performed an investigation of
camouflaged, tactical target detection using the narrow-
band multispectral camera system (Grodewald, 21).

A subtask of the camouflage detection program was the
investigation of techniques for vegetation stress analysis.
One of the prime species of vegetation investigated during
the subtask was corn, because of the extensive ground truth
available for this crop type at the Pennsylvania State
University agricultural test sites utilized during the
camouflage detection-stress analysis study.

Since it was well known that accurate crop yield pre-
diction was a major goal of national agronomists and
economists, RADC personnel took personal initiative to
extend the vegetation stress analysis, developed under the
camouflage detection effort, into a feasibility study in
agricultrual crop yield prediction. The same multispectral
imagery that was collected for the vegetation stress analysis
study was used during the crop yield prediction study. The
crop yield prediction study's associated image processing
effort, data analysis, and documentation were products of
RADC personnel's efforts during after-duty hours. This
documentation serves as a technical report of both the
overall program and its significant findings.

Lt. Gregory B. Pavlin, RADC/IRRE, was the RADC Investigating Scientist
and overall program coordinator. The work described in this report was
performed from March 1975 through March 1976 during the periods that the
RADC Image Processing Facility was operating.

i

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

List of Tables

xi

TABLE OF FIGURES

Executive Summary

Introduction:

Personnel of the Rome Air Development Center (RADC) performed a feasibility study which investigated agricultural crop stress detection, yield measurement, and yield prediction of corn in Central Pennsylvania using a digital, temporally-registered, narrowband multispectral data set. The narrowband multispectral imagery had been collected from an airborne platform over seven geographically-distinct agricultural test sites in Central Pennsylvania during the 1974 corn growth cycle.

This study is significant because the interpretation of the analysis of this unique data set suggests that yield prediction of an agricultural crop, early in the growth cycle, is possible using temporally-registered, specific narrowband multispectral imagery.

Specific subjects investigated in this study include semiautomatic crop identification, stress differentiation, yield measurement, yield prediction (accuracies, models, predictors, timeliness); imagery-weather data relationships; factor analysis results; stepwise and forced regression results (optimal multispectral bands, significant remotely detectable variables); and recommendations.

Background:

In order to operationally predict the yield of agricultural crops with remote sensor systems, the problems of automated crop identification, stress detection, and remote yield measurement must be resolved.

During 1972-1973, the USAF Rome Air Development Center (RADC), located at Griffiss AFB in Rome, New York, investigated the feasibility of crop identification using the narrowband multispectral sensor and related additive color display technology developed by Spectral Data Corporation of Hauppauge, New York. The crop, opium poppies, located in three diversely distinct world environments, was 98% identifiable during all resolvable stages of growth from both low and high altitude sensor platforms. Crop speciation required only one set of four narrowband multispectral filters (Ref: 23).

1

During a limited study in 1973, RADC demonstrated the feasibility of detecting hydrostress in corn fields using the Spectral Data Corporation's (SDC) narrowband multi-spectral sensor system (Ref: 21).

RADC personnel were planning to investigate the feasibility of camouflage detection utilizing the narrowband multipsectral camera system. This effort incorporated the requirement to investigate the feasibility of detecting stressed vegetation. RADC personnel took personal initiative to extend the vegetation stress analysis study to a feasibility study in crop yield prediction. The camouflage detection-vegetation stress analysis study was performed during 1974. The effort's resultant multispectral photography was utilized by RADC personnel to investigate the feasibility of demonstrating crop stress differentiation, yield measurement, and yield prediction.

Ground and Image Truth Programs:

An experiment for demonstrating the feasibility of yield prediction of an agricultural crop in Central Pennsylvania was designed around seven agricultural test sites which included the Pennsylvania State University Agronomy Farm. Corn was the subject crop of the study because of its value as an important food commodity and its complexity with respect to yield prediction through remote sensing technology.

Ground truth, or intelligence obtained on the ground to support the collected imagery, was arranged to be collected by personnel of the Pennsylvania State University Agronomy and Remote-Sensing Departments. This ground truth was arranged in three echelons of detail and included information in the form of weather data, planting conditions, fertilization records, soil mapping, bimonthly site visits for monitoring stress patterns and obtaining new information, and accurate yield measurements in areas of localized stress within the monitored field sites. The ground truth was collected from April to October of 1974.

A ground based radiometry survey was conducted during three phases of the corn growth cycle to obtain the visible-near infrared reflectance spectra of both stressed and unstressed corn. These reflectance spectral signatures of corn were computer processed in order to determine the optimal narrowband multispectral filter set for detection of stresses within corn. The selected filters mounted on the multispectral camera were flown regularly at low-medium altitudes from April to October 1974 over seven test sites

2

of corn varying from 40 to 170 acres in size. High altitude imagery of the field sites was collected during July and October of 1974. The low altitude multispectral imagery acquired at 6000 ft. above ground level (AGL), was supported by flights with other photographic sensors, thus providing image truth to be used in conjunction with the analysis of the multispectral imagery. These additional sensors generated a film data base containing black and white infrared photography and both color and color infrared ektachrome photography (Ref: 20).

Additive Color Analysis and Conclusions:

Upon completion of the multispectral imagery collection, the film was processed under controlled conditions. Additive color analysis was performed with the imagery using an additive color viewer. From this analysis it was determined that the narrowband filters had been successful in maximizing the detection of stress within corn (Figures 1, 2 and 3; Figures attached to Summary). This analysis demonstrated that a narrowband multispectral sensor could detect or differentiate a variety of health or stress levels within corn. The latter statement does not necessarily imply that the cause of the stress can be identified by simple analysis of the additive color displays. To a marginally successful degree, the imagery demonstrated a limited ability to speciate corn from other crops. This ability was demonstrated with low altitude imagery during certain periods of the growth cycle (Figures 4 and 5; Ref: 20,54).

Digital Analysis

Semiautomatic crop identification and corn stress detection were two efforts initiated in parallel, employing the graphics, preprocessing, and logic capability of the RADC DICIFER Image Processing System. Figure 6 illustrates the interactive display consoles of the RADC DICIFER Image Processing System (Ref. 18,32).

The original multispectral imagery was sent to personnel of the RADC Image Processing Facility. The imagery was carefully reviewed to determine which of the seven agricultural test sites were most completely imaged during the corn growth cycle. Due to cloud cover, pilot error, and improper exposure, only the imagery collected over corn field test sites #1, #3 and #7 was usable. The three sites were separated geographically by 10 to 90 miles, occupied different soils, and experienced different degrees of farming management and weather

3

conditions. The photographic data sets of the sites were digitized and registered using the peripheral hardware and preprocessing software of the RADC DICIFER Image Processing System. The digital data set for the three major corn field sites encompassed a total of sixteen minifields where ground truth, stresses, and yield had been carefully measured. Each digital data set was unique in that the narrowband multispectral images (green, red, and near-infrared) were temporally-registered to span the growth cycle of corn at each test site (Figure 7).

Parallel interactive image processing programs were initiated for investigating the feasibility of demonstrating semiautomatic crop identification and corn stress differentiation using the three digital image data sets. The interactive process basically consisted of displaying the images, graphically creating a design or sample set of the classes (i.e., corn, water, various stresses, etc.), evaluating the image features' (tone, texture) ability to differentiate a class, select a decision logic, classify the entire image automatically, check the results and modify procedures if necessary.

Crop Identification:

Recent literature, (Ref. 5,33,45,49) indicates that crop identification accuracies increase using temporally-registered multispectral data. Furthermore, the narrowband multispectral sensor system of Spectral Data Corporation had demonstrated crop identification during the entire growth cycle of opium poppies in three diverse environments of the world, at low and high altitudes, using only one set of narrowband filters (Ref: 23). Though there is no guarantee that such a feat could be duplicated with corn, the remote identification of corn was decided to be only a secondary criteria for selection of the narrowband filters. Consequently, the corn stress detection filters were selected from radiometry of corn during the summer of 1973 and during three phases of the corn growth cycle in Central Pennsylvania during 1974. However, considering that an ideal futuristic multispectral sensor system dedicated to crop yield prediction would have narrow bandpass detection which both speciated crops and detected stresses simultaneously, the digital data set for stress detection was image processed to assess its potential for the semiautomatic speciation of corn.

Using the RADC interactive image processing capability reviewed in the Digital Processing Section, high altitude multispectral film imagery collected at 50,000 ft AGL was digitized, registered, and preprocessed. At least thirty different methods of class sampling and decision logic were evaluated to digitally speciate corn from other crops. The best decision images are presented in Figures 8, 9, 10, and 11 and demonstrate about 70% overall classification accuracy. Low altitude imagery displayed better results using semi-automatic processing techniques.

Crop Identification Commentary:

The overall classification results are better than expected considering the objectives of the filters were to detect and map stresses within corn and not to speciate corn throughout the growth cycle. The literature indicates that multispectral, temporally-registered data improves classification accuracy.

The conclusions of this division of the study venture beyond mere speculation that virtually 100% semiautomatic crop speciation is possible using temporally-registered, narrowband multispectral imagery. This identification accuracy could minimally occur during the period of the crop's growth cycle when remote sensing data and yield associations were greatest (refer to yield prediction section). Conceivably, the digital crop speciation could be implemented any time during the entire growth cycle (i.e., digital opium poppy crop detection). What remains to be determined is: (1) which set of multispectral filters speciate one or more crops during key phases of the growth cycle; (2) would the bands selected for crop identification satisfy the requirements in the yield prediction model for filters which distinguished healthy from stressed vegetation.

Stress Detection in Corn:

Employing a philosophy which assumed the fields of corn at Sites One, Three, and Seven had been semiautomatically identified, the image processing analysis was directed to the demonstration of stress differentiation within the field. The basic image analysis procedure for stress differentiation consisted of displaying the digitally-registered multispectral images of October for the given site, creating a design set of the monitored stress areas where yield measurements had been collected, and stress mapping the fields using computerized decision algorithms. The decision software utilized both the

5

unique tone and simple texture features sampled in the stress area design set for classifying the entire scene. Figure 7 illustrates a typical digital image data set of an arbitrary corn site with the field and design set stressed areas' boundaries in registration.

Approximately twenty different methods for digital stress mapping were attempted. The best stress differentiation results occurred when Fisher discriminant logic was employed with a design set sample size selected with the Foley Criteria (Ref: 17).

Computerized image processing could differentiate and map stress patterns of corn using the temporally-registered multispectral data set. Qualitative analysis indicates that there was a correlation between vegetation patterns and soil patterns imaged in April over the respective sites.

Yield Measurement:

Prior to investigating the utility of the digital data for yield prediction, a study was performed investigating the potential of the imagery for measuring yield prior to harvest.

The October 1974 multispectral imagery for Sites One, Three and Seven were semiautomatically stress-mapped using the procedures outlined in the preceding section on stress mapping. Since the small box areas of the design set for stress mapping correspond to the exact geographic location where yield had been measured (Figure 7), the digital stress map of the site was also a digital map of corn yield as a function of stress. Accuracies for yield measurement ranged from 96 to 98% for the three large fields of corn. This figure of accuracy was obtained by integrating the yield value assigned to each picture element (pixel) of the digital stress map over the entire field and comparing this result to the independently measured average yield of the field.

Figures 12, 13 and 14 illustrate examples of computer-generated stress maps of the three field sites. These decision images were generated from the digitally-registered green, red, and infrared images of the multispectral imagery, the stress area design set, and Fisher discriminant logic.

Yield Measurement Commentary:

Semiautomatic stress differentiation could be replaced by an automatic classifier consisting of an unsupervised clustering or iterative discriminant logic operating on only

6

the picture elements that had been previously identified as corn. Yield measurement would then be a semiautomatic process which linked the ground truth measurements per stress to the registered multispectral image stress map. Conclusions of the yield prediction study suggest that conjecture could be extended to include a completely auto-mated yield measuring system using only remotely sensed data.

Yield Prediction:

The hypothesis that agricultural crop yield prediction can be accomplished with a remote sensor has developed from both physical theory and field experiments. The literature (Ref: 5,10,11,29,30,33,43,45) identifies several researchers who have linked the measure of a crop's photosynthetic potential or yield to such parameters as leaf area, leaf to soil ratios, and brightness coefficients (albedo); parameters measurable from a remote sensing device!

The dual success of both semiautomatically measuring the yield and differentiating stresses within corn imaged in a multispectral data set of each of three geographically dis-tinct corn field sites with a 96+% accuracy gave impetus for investigating the feasibility of yield prediction using the multispectral, temporally-registered data set.

A measurable image feature had to be recognized as common to each site's six multispectral imagery data sets (April - October), before any hope for relating remotely sensed data to yield could be established. This observation was attributed to the imagery's tonal and textural features being a function of film processing, atmospheric conditions, illumination, digitization, etc., along with the health of the crop. The image-derived measure was identified to be the variance in grey level value from the mean grey level of all points that were classified to be corn. Figure 15 illustrates the latter concept. The model of Figure 15 is acceptable if the imagery has been processed such that points labelled corn are located in the linear portion of the film's density/exposure curve. Discounting other unmentioned technicalities, this unique image measure (DIFX) was selected to be the numeric difference between the average grey level of the sampled stress areas and the average grey level for the entire field. The variance fluctuation over time of the remotely-sensed multispectral data was hypothesized, prior to this study, to be a possible yield predictor variate. To test the hypothesis, these measures, the latter difference feature (DIFX); the average grey level for the entire corn field (per band per month) (GLTX); and the average grey value for each stressed region

7

(YSAX) where yield was precisely measured (per band per month) (Figure 7), were analyzed along with weather data and ground truth data to determine the significant independent variables associated with yield. Extensive principal component analysis and factor analysis were performed with the data. The study was completed with an extensive stepwise and forced regression analysis program using the computers of the RADC Information Processing Facility. The significant relationships and conclusions of this study are presented in the conclusion section of this report.

Conclusions of the Feasibility Study in Crop Yield Prediction:

1) The feasibility of timely (as early as pollination phase) and accurate prediction of agricultural crop yield (87% of actual crop yield variance accounted for with 90% accuracy), for fourteen small corn fields in Central Pennsylvania, was demonstrated utilizing a yield prediction equation that was solely dependent upon remotely-sensed data; specifically, a digital, temporally-registered, multispectral image data set.

2) When compared to the variables defining the time-variant, multispectral image data, the weather variables were observed to be insignificant predictors of agricultural crop yield.

3) Though multispectral image data sets, collected on 29 April, 2 July, 16 July, 19 August, 4 September, and 8 October 1974, were analyzed to determine the feasibility of crop yield prediction, the variables defining the imagery collected during mid-July and mid-August were prime predictors of agricultural crop yield. Optimal multispectral imaging periods for accurate yield prediction of the same Pennsylvania test sites may actually occupy some subset of the latter time interval (Regressions 1-6,7,8,13,15).

4) Though two temporally-independent multispectral image data sets were found to be adequate for timely and accurate yield prediction, the analysis of three or more image data sets, spanning from early July to early September, did increase the percent of yield data variance accounted for to nearly 100% with a 90% accuracy (Regressions 15-19 vs. 7,8).

5) The yield analysis from two or more bands of multispectral imagery, registered over time, provides better prediction results than when only one band's imagery is utilized. The optimal two-band yield predictor combination was the green/infrared image record combination. Highest accuracies in yield prediction occurred when image data from all three bands was analyzed (Regressions 1-6 vs. Regressions 7,8,11,12).

8

6) The average density of the digital multispectral images of the three large corn agricultural test sites, defined by the variable "GLTX" in Figure 7, was highly correlated to weather data. In a regression equation for yield prediction, GLTX could be substituted for weather variables with no decrease in the prediction accuracies.

7) Some film variables were found to be better yield predictors than others. Surprisingly, the film variable "DIFX", created to empirically remove the effects of variability in image density due to film processing and exposure so that monthly multispectral imagery data might be linked by some commonality feature, provided less yield prediction information than the variable defining the actual digital, nonadjusted grey level (YSAX). This discrepancy could not be explained but should be resolved in a follow-on study (Regressions 1,3,5,7,11,13 vs. 2,4,6,8,12,14).

Recommendations:

A program of significant magnitude, spanning an estimated three years, should be developed to verify the demonstrated feasibility of agricultural yield prediction with temporally-registered narrowband multispectral imagery. The proposed study should be directed to demonstrating high accuracies in digital crop identification and yield prediction using solely, remotely-sensed data.

The first year of the program should be devoted to collecting in situ spectroradiometric data of important agricultural crops and typical background materials throughout the respective phases of the growth cycle. The analysis of the radiometric data would determine the narrowband filters or bandpasses of detectors for speciating the selected crop(s) during the growth cycle.

During the second year a calibrated narrowband multispectral sensor, either photographic or electro-optical, operating at the selected bandpasses, should be flown at low and high altitudes over geographically and environmentally diverse areas of the United States. Adequate ground and image truth should supplement the multispectral imagery collection effort.

The narrowband multispectral imagery would then be digitally analyzed with an image processing computer system dedicated to handle the large volume of data collected during the study. Automatic crop identification and yield prediction would be investigated with image processing techniques, factor analysis, and regression analysis during the third year of the study.

9

The proposed effort would ultimately establish accuracies that might be obtainable with respect to automated crop identification and yield prediction, using a temporally-registered, narrowband multispectral data set. Additionally, the key phases of the growth cycle for yield prediction and/or crop identification for a major economic crop could be determined. System effectiveness over geographically-distinct environments could be measured. Finally, the cost effective investigation would provide valuable information for specifying characteristics of an operational satellite system dedicated to yield prediction and crop identification.
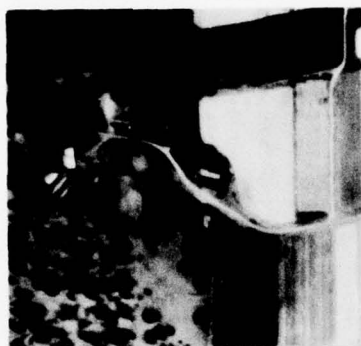
1 May 1974 5500 Ft AGL   2 July 1974 6000 Ft AGL

16 July 1974 10000 Ft AGL   19 August 1974 10000 Ft AGL
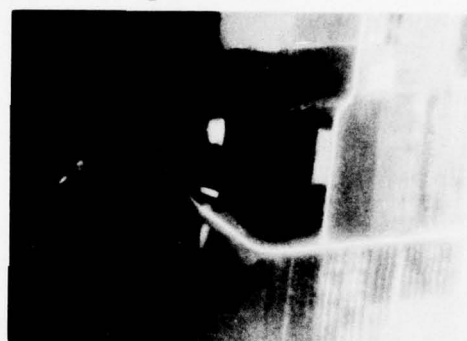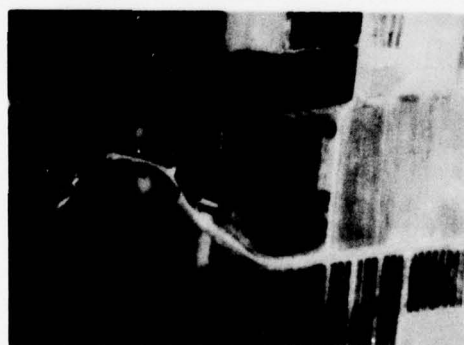
4 September 1974 6000 Ft AGL   8 October 1974 10000 Ft AGL

Figure 1: Six sequential multispectral additive color images of Site One: Multispectral black & white photography collected on the respective dates generated the additive color renditions on the SDC additive color viewer when the green image was color projected as green, the red image as red, and the infrared image as blue. Note the stress patterns within the field which are most noticeable in the 2 July, 16 July and 4 September images.

11
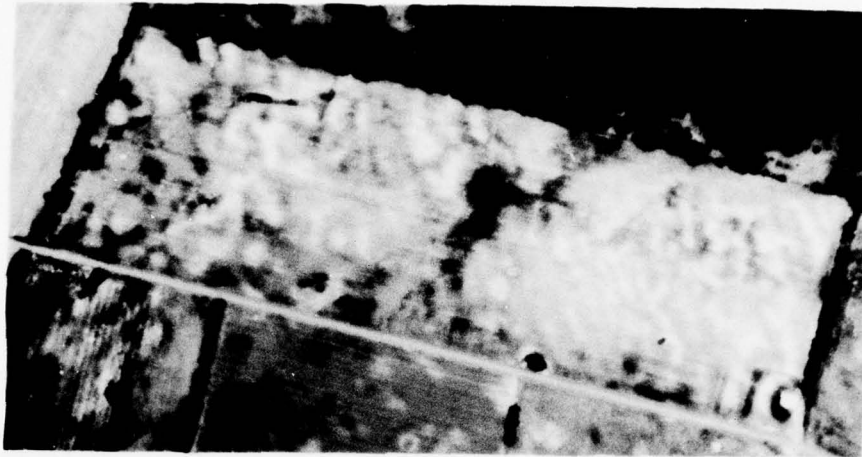
Figure 2: Multispectral additive color rendition of Site Three; Multispectral black & white photography collected on 4 September 1974 from an altitude of 6000 ft(AGL) generates the additive color rendition on the SDC additive color viewer when the green image is color projected as green, the red image as red, and the infrared image as blue. Notice the stress patterns within the field that are apparently enhanced.
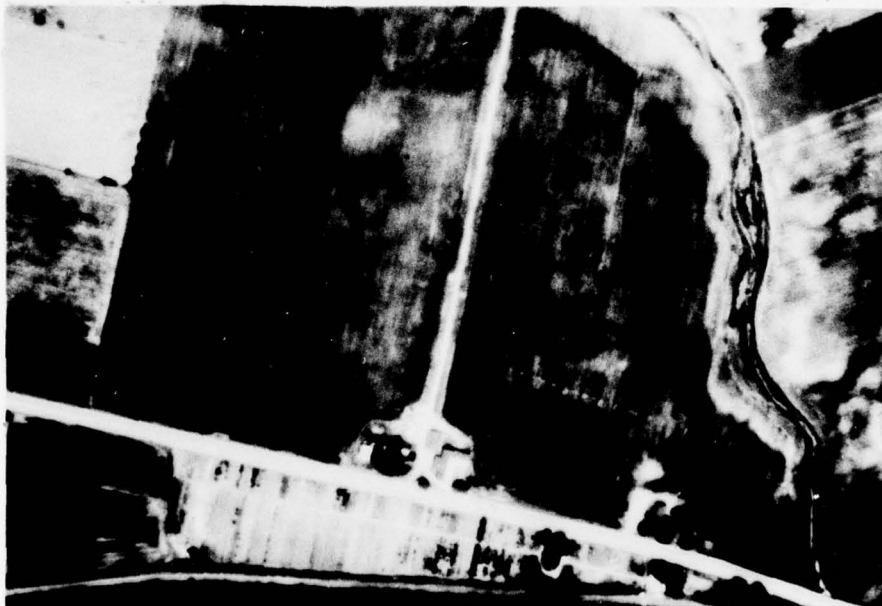
Figure 3: Multispectral additive color rendition of Site Seven; Multispectral black & white photography collected on 2 July 1974 from an altitude of 6000 ft(AGL) generates the additive color rendition on the SDC additive color viewer when the green image is color projected as green, the red image as red, and infrared image as blue.

12
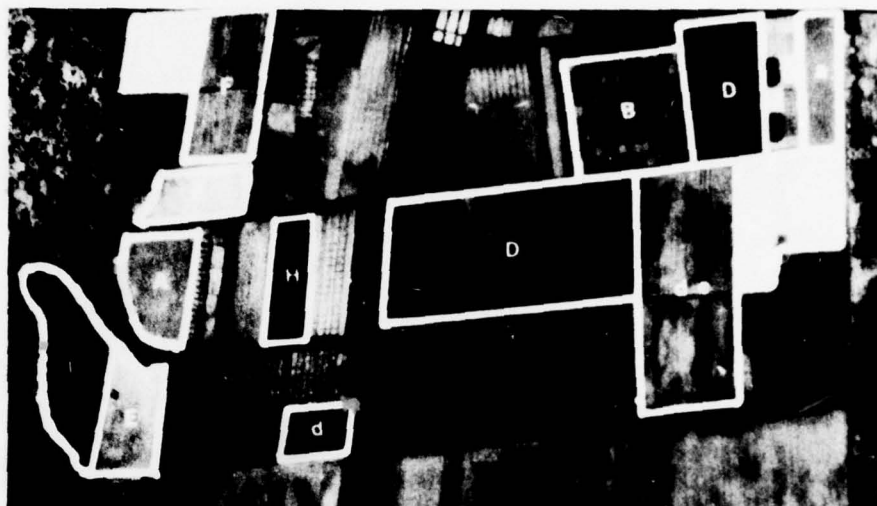
Figure 4. Annotated image truth photograph of the same area as shown in Figure 5.

FIELDS	CROP AND DESCRIPTION
A	This is a normal corn field.
a	This is a "No Till" corn field, it has a high percentage of weed cover at ground level.
B	This is a field containing ordered replications of corn, oats, crown vetch, soybeans, and wheat. The individual plots are approximately 10' x 20'.
D	Oat stubble, the oats in this field were harvested before the photograph was taken.
E	Alfalfa
de	This field contains a combined planting of oats and alfalfa, it must be considered an experimental field.
d	Headed oats.
F	Wheat
G	Potatoes
H	This field contains a combined planting of wheat and timothy, it must be considered an experimental field.
I	Pasture grass and timothy.

Note: Some of the other experimental fields do contain corn, however, the plots are small and not in the normal stage of growth for this time of year.

Upon examination of the ground truth photograph in Figure 4, along with the multispectral rendition in Figure 5, it can be seen that no other crop at the Agronomy Farm, other than corn, displays exactly the same color. When examined on the additive color viewer screen, the most similar color to corn was that shown in the field marked "4a" in Figure 4, which is an experimental field which combined the planting of alfalfa and oats.

Figure 5: Photographic reproduction of an additive color viewer screen presentation of part of the Pennsylvania State University Agronomy Farm, illustrating the speciation of corn from other crops. Corn fields and test plots exhibit a characteristic brown-red color. The ground truth for this area is shown in Figure 4. Mission 00074-47, frame 4), 16 July 1974, Altitude 10,000 feet ASL. Rendition: The 541-581 nm (green) band is imaged as green, the 620-680 nm (red) band as blue, the 810-955 nm (infrared) band as red.

14

Figure 6: The Rome Air Development Center(RADC) Image Processing Facility employs the interactive man/machine interface for effective image processing. The DICIFER Image Processing System, described in the following summary, was a key tool utilized during this feasibility study of crop yield prediction and remote crop identification.

RADC DICIFER Image Processing System

RADC DICIFER IMAGE PROCESSING SYSTEM

Overall Description:

The DICIFER Image Processing System is a flexible interactive system which provides the analyst with the capability of fully exploiting multisensor data in photographic, analog, or digital format. The system provides the user with numerous options to format, edit, manipulate, analyze, and extract features to be used in the application of pattern classification techniques.

Processing Hardware:

A PDP-11/20 computer with 28K memory and hardware multiply and divide serves as the analysis system. Peripherals include a card reader, line printer, Tektronix 4010 CRT and keyboard (with hard copy), a 256K word fixed head system disk, DECtape and 10M word disk pack.

Input Hardware:

1. Photographic input is accomplished by a SDS IV vidicon scanner capable of handling film formats from 70mm to 9 inches. Digitization is performed by the SDS 108 system which formats the image into 512 lines with 512 pixels per line.

2. Line scan data is converted by a 7-channel A/D station. The sampling rate may be varied from 10KHz to 300KHz. A maximum of 2000 pixels are produced per line of data.

3. Both 7 and 9-channel digital tape units are available for digital data input.

Output Hardware:

1. A SDS 800 32 Color TV monitor with cursor is available for interactive data analysis. A black and white monitor is also available.

2. Hardcopy output of data is handled by a digital color printer capable of printing, on five inch film, 1024 pixels per line with 64 gray levels per pixel.

Software:

1. DOS V09 is available on the PDP-11/20. The system includes a Fortran IV compiler and MACRO capability.

2. The DICIFER image processing system provides interactive data analysis software capable of manipulating N-dimensional data. Many data editing, manipulation, and display options are included. Tone feature extraction capability along with first and second order texture feature discriminants are a part of the system capability. Both Boolean and Fischer discriminant classification logic are implemented on the system.

15

TEMPORALLY REGISTERED MULTISPECTRAL DIGITAL DATA
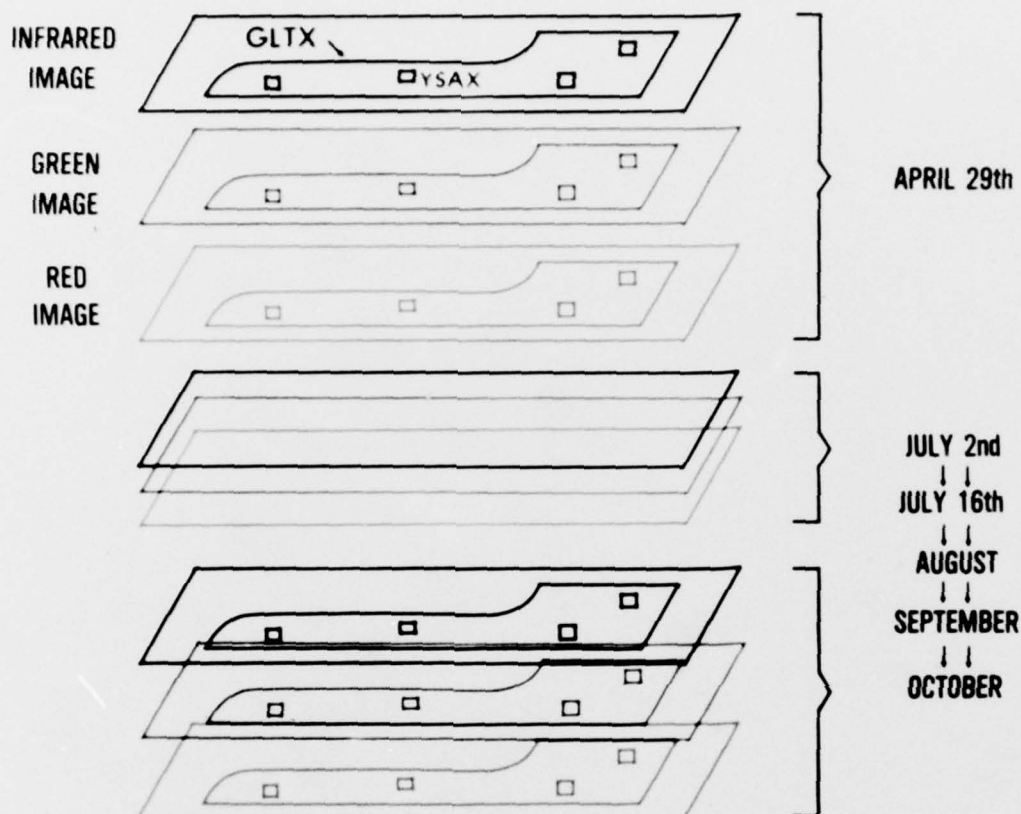
Figure   7:   Illustration of the digital, temporally-regi-
                 stered, multispectral image data set assembled
and utilized in the semiautomatic crop identification, stress
differentiation, and yield prediction studies.   The image-
extracted variables used in both the multivariate and regres-
sion analyses' studies may be more fully understood upon
referencing this figure:

- Image variable "GLTX" equated to the average density of
  all pixels contained within the site boundaries per band
  per month.
- Image variable "YSAX" equated to the average density of
  all pixels contained within each minifield's area file
  per band per month.
- Image variable "DIFX" equated to the numeric difference
  between "YSAX" and "GLTX" for each minifield's area
  file per band per month.

16

Figure 8: High Altitude (50,000 ft. AGL) digital image of Field Sites 6 and 7: (red band) with annotated scene classification: C - corn; T - trees; W - water; O - other.
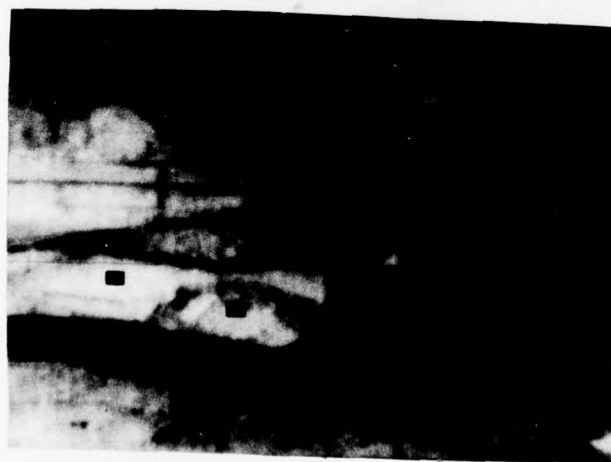
Figure 9: High Altitude (50,000 ft. AGL) digital image of Field Sites 6 and 7: (infrared band) with small black boxes mapping where corn, trees, and water classes were sampled for performing semiautomatic crop identification.

17

Figure 10: High Altitude Crop Identification Decision Image
Generated from Fisher logic criteria applied to tone features
from the green, red, and infrared images along with six tex-
ture features extracted from each multispectral image. Ref-
erence Figure 8 for comparison. Red-Corn; Yellow-Trees;
Light Blue-Water; Purple-Rejects. Accuracy--80%

Figure 11: High Altitude Crop Identification Decision Image
Generated from Fisher logic criteria applied to only tex-
ture features extracted from red and infrared multispectral
images. Reference Figure 8 for comparison. Red-Corn; Yellow-
Trees; Blue-Water and Rejects. Est. Accuracy-19+%

18

Figure 12: Digital Stress/Yield Map of Field Site 1 generated with Fisher
logic criteria and the grey level feature extracted from the green, red,
and infrared images of all of the temporally registered (April - October)
multispectral images. Accuracy of the yield measurement: 100%

Note: Each color corresponds to a unique stress, positive or negative,
mapped within the field. Since yield is a function of the stress, the
digital image is also a map of yield within the field.
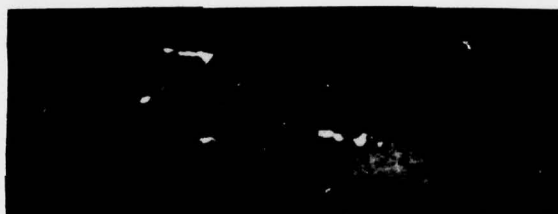
Figure 13: Digital Stress/Yield Map of Field Site 3 generated with Fisher
logic criteria and a grey level variance feature extracted from the green,
red, and infrared images of the October multispectral imagery. Accuracy
of the yield measurement: 90+%
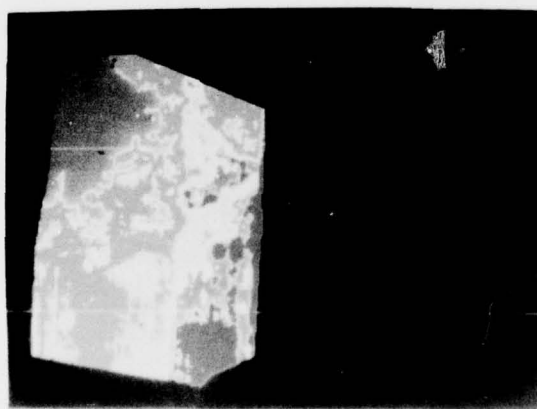
Figure 14: Digital Stress/Yield Map Field Site 7 generated with Fisher
logic criteria and the grey level feature extracted from the green image
of September. Accuracy of the yield measurement: 90%

1974 CORN GROWTH CYCLE
CENTRAL PENNSYLVANIA CORN
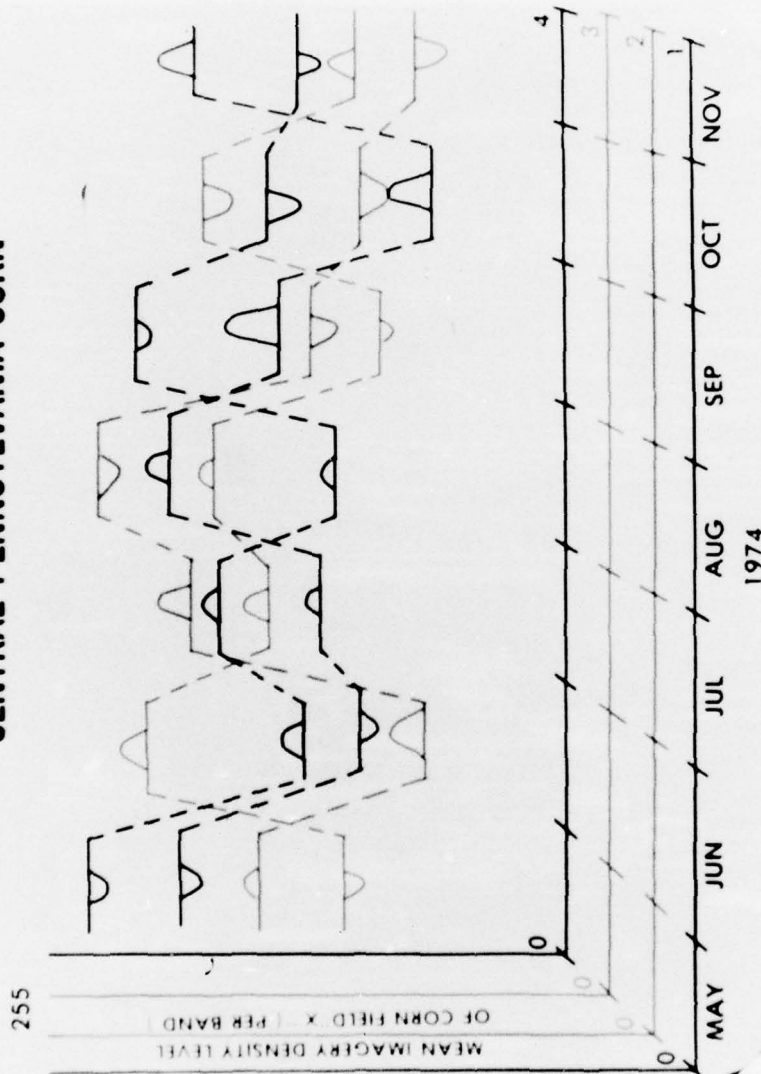
Figure 15: Concept illustration of the yield prediction model investigated during this feasibility study. The study utilized narrowband multispectral imagery which was digitized and temporally registered. The photographic information that was common to the imaged scene over the entire growth cycle was the noise level of the film. Since the film was processed such that the mean density level of imaged subject material, corn, occupied the linear region of the film density/exposure curve, the health of small areas with the corn field would be characterized by the variance in density that was greater or less than the mean density level. It was the analysis of the multispectral fluctuation of density variance over time that established a significant relationship with the yield of the crop.

20

Figure 16: Regressions 1,3,5 and 2,4,6 illustrate the utility of each separate multispectral band's independent contribution to predicting yield using YSAX and DIFX film-predictor variables respectively(Figure 7). The graphs suggest that an ideal period for yield prediction may occur during the mid-July to August period. R^2 is both a measure of the total yield data variance accounted for by the regression-analyzed film variables and of how well crop yield can be predicted. FFR, Forced Forward Regression, and FRR, Forced Reverse Regression, were two types of regressions performed on the data. The titles imply that either May-October or October - May data were forced into the yield prediction regression equation.

21
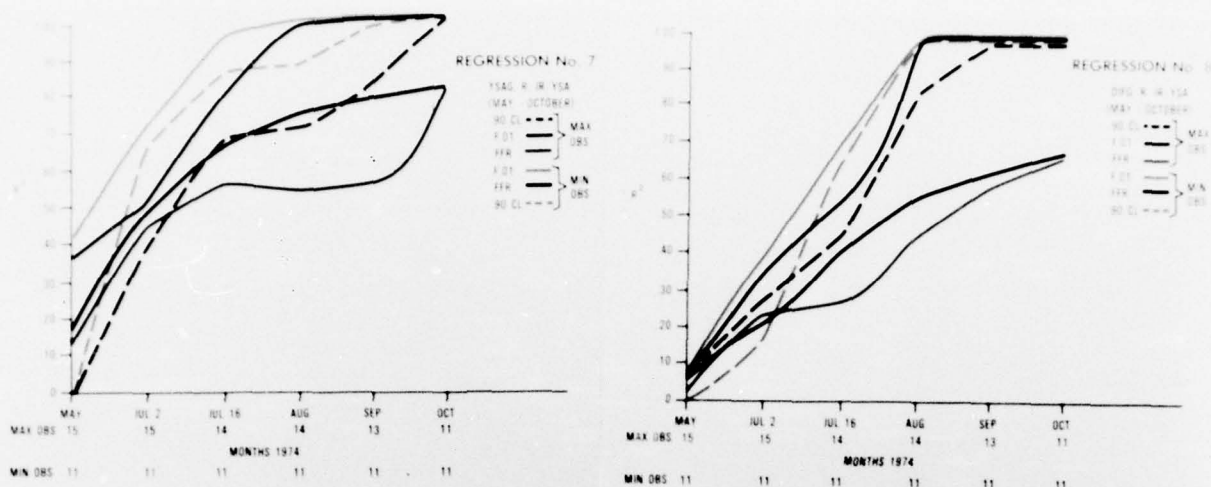
Figure 17: Regressions 7 & 8 illustrate the utility of using all three multispectral images' data for predicting yield using the YSAX and DIFX film-predictor variables respectively(Figure 7). Beginning with May's film data, each month's film data was consecutively added to the regression data set to determine how early in the growth cycle could yield be predicted with accuracy.
Definitions:
MAX & MIN OBS: Sample size used in the regression analysis-indicates the maximum number of corn fields per month which had complete image coverage from May to the subject month. Only 11 fields had coverage from May to October. Note: Separate regressions were performed for both MAX & MIN sample sets.
FFR: Forced Forward Regression(See Figure 16 for explanation)
F.01: Stepwise regression at a minimum confidence(accuracy)level: useful for determining which variables were significant for yield prediction
0.90CL: Stepwise regression whose R^2 value gives a measure of how much crop yield is predictable with a 90% accuracy.
R^2: Total yield data variance that could be accounted for in the regression analysis
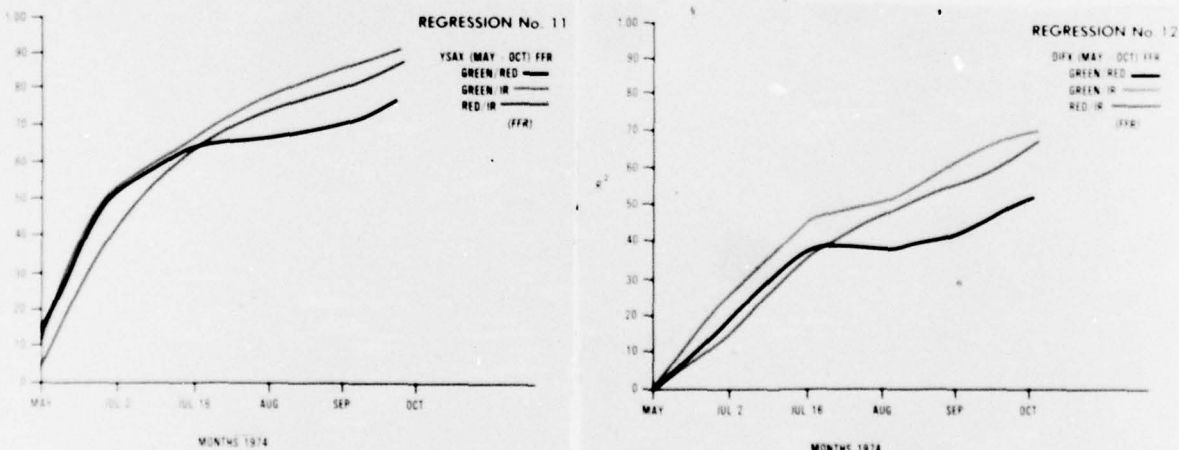
Figure 18: Regressions 11 & 12 illustrate the advantages of certain two band combinations of multispectral film data for predicting crop yield. Graphs plot the results of a forced forward regression.

22
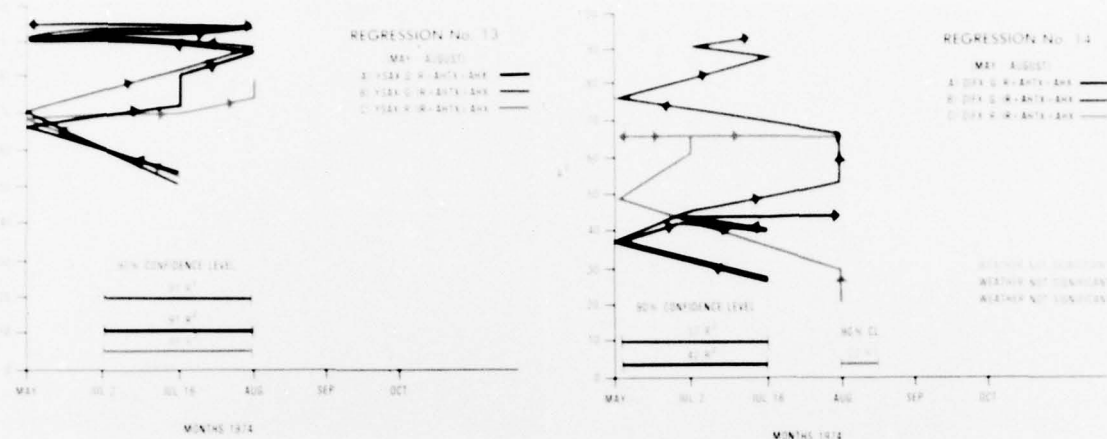
Figure 19: Stepwise Regressions 13 & 14 illustrate the advantages of certain two band combinations of multispectral film data when regressed with certain weather data for yield prediction. The arrows on the graphs reflect the month's data which was considered most significant for yield prediction. YSAX variable was a better predictor than DIFX. Optimal dates for yield prediction span from 2 July through mid-August. Weather was significant when used in conjunction with the YSAX variable.
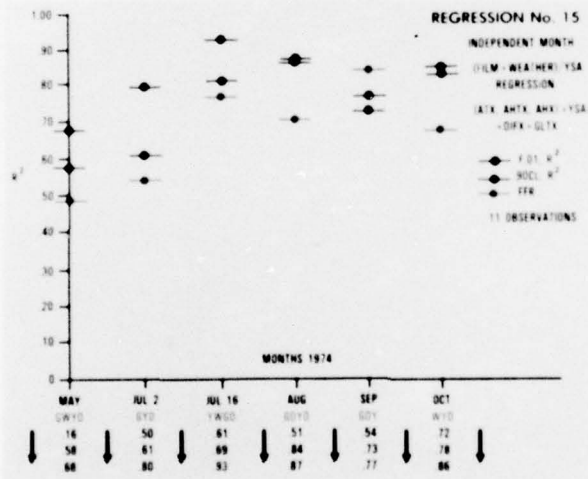
Figure 20: Regression 15 illustrates which month's film and weather data are most significant for yield prediction. Three regressions were performed for each individual month's data: Stepwise regressions at both low and high accuracies (F.01 and 0.90CL, respectively); Forced forward regression. July 16 and August data are most significant for yield prediction. For variable definitions reference Figure17.
Special Definitions: ATX-Average Temperature/month; AHX-Average Humidity/month; AHTX-Average High Temperature/month.

23
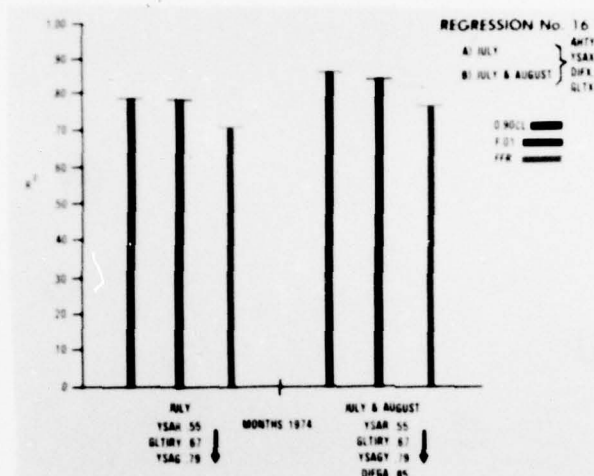
Figure 21: Regression 16 illustrates the significant increase in yield
data variance accounted for when July's film and single weather variable
is combined with similar August data in a regression equation for
crop yield prediction.
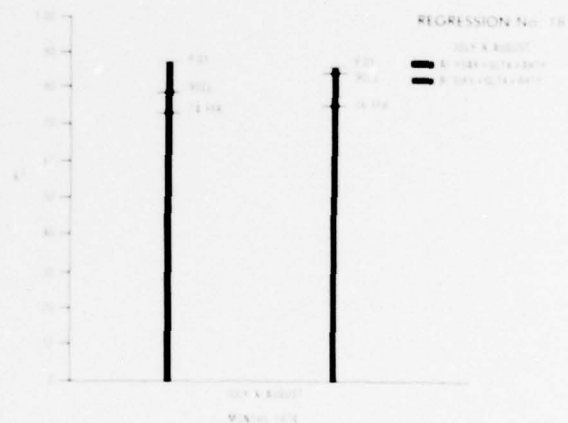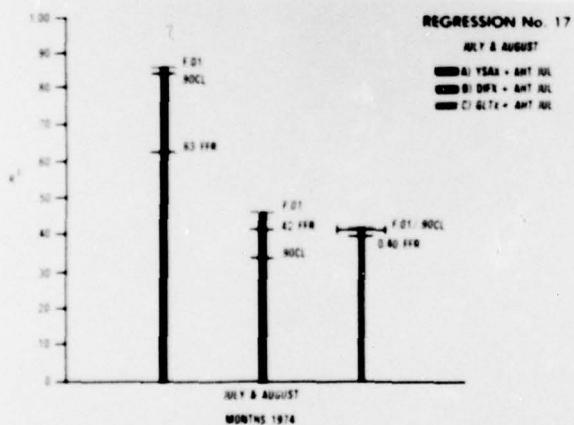
Figure 22: Regressions 17 & 18 illustrate the amount of yield data
variance accounted for when specific film/weather data combinations
are employed in the regression analysis. Three types of regressions
were performed: Stepwise(F.01), Stepwise(0.90CL or 90% accuracy); and
forced forward regression(FFR), (See definitions, Figure 17).

24

Literature Review: Agricultural Crop Yield Prediction

Introduction

Scientists have been attempting to discover significant factors associated with the yield of agricultural crops for many centuries. If such factors could be identified and measured, predicting the yield of crops early in the growth cycle might be feasible. Over the years, such factors as weather, soil moisture, sunlight, soil mineralization, fertilization, farm management practices, and others have been found to be highly correlated to crop yield. With the advent of remote sensing from aircraft or satellite platforms, yield prediction using remotely sensed data became an objective for agricultural scientists. This literature review provides a general overview of the research performed in the interest of advancing crop yield prediction technology.

Agricultural Crop Yield Factors

There seems to be a general agreement by both agrometeorologists of the USSR and the USA that the maturation process and health of agricultural crops depends upon the dynamics of soil moisture supply, weather, and cultivation procedures (Chirkov(10),(2), McKeever(36)). Chirkov(10) identifies air, light, heat, moisture, and soil fertility as the general life factors that determine plant maturation. There have been

numerous studies of each of the life factor's particular
relationship to yield. For example, McKeever(36) performed
a thorough study of soil/crop yield relationships for various
major agricultural crops in Pennsylvania. His study developed
an index whereby, given the soil variety and location in
Pennsylvania, a baseline estimate of yield could be made.
This estimate would be adjusted as a function of climatic
conditions, farm management practices, soil slope, and erosion
classification. The Wyoming Agricultural Experiment Station(2)
performed an analysis of the effects of precipitation on
vegetation and associated yields. Farmers in the United States,
with the help of various county agricultural extension ser-
vices, are more aware of the utility of proper fertilization,
acidity control, insect control, weed control, planting
practices, irrigation, and other farm management practices.
All of these studies' specific factors can be condensed to the
life factors identified by Chirkov(10).

Yield Prediction Methodologies

Though life factors and yield factors have been identified,
a plan to nationally survey and measure such factors along
with agricultural management practices is impractical. Such
an operation would be expensive, time consuming, and ultimately
untimely in its ability to predict the yields of crops on a

26

national or international scale. Yield prediction programs
in both the USSR and USA attempt to surmount this problem by
employing sampling techniques and statistical inference
theory (Chirkov(10), MacDonald(33), O'Dell(40), Ecosystems(50)).

Chirkov(10) describes the mechanics of crop yield pre-
diction in the Soviet Union. The methodology involves the
remote monitoring of crops, extensive networks of agricultural
stations, ground truth, and computer processing. Specifically,
the Soviet crop forecasting procedures might be divided into
three phases: (1) agrometeorological data is collected
through a network of stations which perform airborne and
surface examination of crops, controls the data, and processes
the data with computers; (2) an estimation of initial crop con-
ditions is generated from information collected at the time
of the forecast; (3) using prognostic regression equations
based on the yield-related variables identified in Chirkov's
report, yield is calculated and readjusted by intensity factors
generated from apriori knowledge of the regional crop. This
approach reportedly achieved a 90-92% accuracy over a ten year
period. Chirkov fails to state at what point in the growth
cycle such accuracies in yield prediction could be obtained.
Chirkov commented that there are other approaches to yield
prediction in the USSR which utilize complex mathematical
models to describe the crop maturation process. The author
sites two reasons for the failure of such models: (1) the
observations on the ground required for the model cannot be

27

handled by a routine network of agrometeorological stations;
(2) the models require too much data to be processed.

The United States Department of Agriculture (USDA)'s
Crop Reporting Service uses a national statistical sampling
for the development of crop production estimates which serve
as an input to crop production forecasts (Earth Satellite
Corporation(1)). In a 1974 paper, Ecosystems Inc. (50)
reported that the United States government spent close to
$40 million a year to survey the United States crops for
predicting and measuring the total crop yields. The article
noted that for the years of 1969-1970, the USDA technique was
achieving an absolute forecasting accuracy of 8%. Ecosystems
(50) elaborated on the USDA approach to yield prediction.
Initially, questionnaires are sent annually to farmers on a
national scale for the purpose of estimating the acreage of
crops being planted. Conditional yield estimates are made
in July and continue through October. The information is
reportedly released to the public during the same months
without a great deal of lag time. The yield estimates are
generated by the USDA through a regression analysis dedicated
to yield prediction. The independent variables incorporated
into the regression equation include: (1) a measure of the
condition of the crop at the time of sampling; (2) inches
of rainfall during the two months prior to the statistical
sampling; (3) forecasted rainfall during the two months
following sampling; (4) calendar time elapsed since the last

28

sampling interval. The regression equation coefficients are derived from analysis of national agricultural records collected during the last fifteen years.

Despite the current yield prediction programs in the United States and USSR, scientists are still attempting to improve the yield prediction accuracies through various modeling techniques. Bryson and Starr(48) discuss a unique methodology for predicting the yield of agricultural crops on a large scale. Initially, having identified the important nonlinear response of crop yield to weather, Bryson and Starr propose to model climate so that mean monthly temperature and total monthly precipitation can be predicted. The latter model is incorporated with another model which predicts the estimated acreage of crops using socioeconomic predictors. By integrating the two models, the authors intend to predict the yield of crops for various periods of time. The authors commented that their approach to yield prediction of crops was a consequence of their discovering that too many yield prediction models being developed in the United States simply required too much ground truth and data processing.

The science of agricultural crop yield prediction is transitioning to utilizing remote sensing technology. To date, the United States and other countries have attempted to predict the yield of crops through elaborate remote sensing ground truth sampling programs (MacDonald(33), O'Dell(40)).

29

With the development of the National Aeronautics and Space

Administration (NASA)'s Earth Resources Satellites Landsat I

and II (formerly ERTS-1 and ERTS-II), scientists from all

over the world have hoped to utilize remote sensing tech-

nology to solve the problems of crop yield prediction. The

rationale behind this trend in scientific thought may be

based on what information a surveillant earth resources

satellite provides. The basic information consists of an

image of a large geographic area of the earth. If crops could

be identified in the image with high accuracy, an estimation

of crop acreage could be made. The latter acreage estimates

integrated with information acquired from proper monitoring

of climatic conditions during the growth cycle, might make it

feasible to estimate the yield of a crop early in the growth

cycle. Of course it would be even more ideal if the yield

of the crop could be predicted by simply analyzing the image

data of the satellite sensor.

The Scientific Rationale for Remote Yield Prediction

Remote sensing experts were keen on demonstrating three

capabilities with the launching of the earth resources

satellites; specifically, remote crop identification, stress

detection, and yield prediction. Obviously, there needed to

be some scientific rationale for believing that remotely

sensed data could be related to yield.

30

Larin and Lebedev(30), Rachkulik and Sitnikova(43), Kumar and Silva(29), and MacDonald(33) have directly or indirectly established the relationship of brightness or albedo of a crop's canopy to yield. Larin and Lebedev(30) state axiomatically that high yield figures are determined by the efficiency of the photosynthetic process. Efficient photosynthetic processes result when the life factors described by Chirkov(10) are mixed optimally. In the case of corn, Chirkov specifically identifies the following as factors affecting yield: soil moisture in the spring; precipitation during the intermediate phases of the plant's maturation cycle; average air temperature; supply of moisture averaged over the maturation phase of the corn plant; indications of crop structure (i.e., plant height, number of cobs/plant); humidity deficiency; area of leaf surface at flowering per hectacre. Using these factors and a few others, along with quadratic regression equations, Chirkov is able to predict the yield of corn for oblast size regions of the Soviet Union with an accuracy of 83%. Unfortunately, there exists an inherent problem with yield prediction models requiring ground truth measurements; specifically, the ground truth measurements.

Chirkov(10) observed from his studies of specific life factors that the yield of corn was highly correlated to the area of the leafy surface, plant height, moisture in the 0-50cm region of the soil, precipitation from panicle formation to

31

waxy ripeness stage, and the average air temperature for the same precipitation period. He also discovered that the measure of leafy area seemed to be a good measure that incorporated other life factors, including the influence of meteorological conditions.

Larin and Lebedev(30) noted that a good measure of the efficiency of a photosynthetic process is the size of the leaf area; i.e., the larger the leaf, then the more efficient is the photosynthetic process. This thought was studied and expanded by several scientists who determined that a measure of photosynthetic efficiency is actually the leaf to soil ratio (Chirkov(10), Colwell(11), Lebedev(30), Rachkulik(43), Kumar(29), Roberts(45), and Bauer(5). Rachkulik(43) determined several important relationships linking leaf to soil ratio or leaf area index with the "brightness coefficient" or the radiometrically sensed albedo of the plant canopy. He concluded that when a condition of high contrast existed between soil and dense plant cover albedo, a more pronounced relationship was observed between the brightness coefficient of the soil-vegetation system and the quantity of vegetation above the ground. Significantly, the relationship between the leaf area and the brightness coefficient makes it possible to monitor variations of leaf area with respect to time using airborne observatories; and hence, to compute the photosynthetic potential of crops. Rachkulik concluded that "airborne

32

observations of the brightness coefficients in the various

phases of the growth cycle of vegetation make it possible

to detect deviations from optimum.....agronomical and biological

crop yield may be characterized by maximum leaf area and

surface vegetation mass. Since the reflectance properties

of plants are determined by photometry of the leaf area, the

brightness coefficient measured at the time of maximum leaf

area will provide a measure of crop yield."

Colwell and Suits(11) reported very significant conclusions

concerning the relationship between an agricultural crop's

yield and the temporally-monitored albedo of the crop's

vegetative canopy. Initially in the study, a preliminary

model describing the growth process and associated yield of

wheat had been developed. Using a model of vegetative canopy

reflectance and the modeled growth process, the wheat canopy

reflectance was calculated. Colwell and Suits then compared

the predicted wheat canopy reflectance to the modeled growth

and yield characteristics of the wheat as a function of time

so that any relationships might be determined. The study makes

the important conclusion that the yield of wheat might be

determined by analysis of the temporal fluctuation of wheat

canopy reflectance. Though the latter premise had been

supported by Rachulkik's(43) radiometry survey, the concept

had yet to be proven using an imaging system. Both Colwell

33

and Rachulkik's findings are supported by the experimental studies of Kumar and Silva(29), Ausmus(3), Gausman(19), MacDonald(33), and the literature reviews of Bauer(5) and Roberts(45).

As a consequence of the studies linking a vegetative canopy's brightness coefficient to its yield potential, the problem of detecting stressed vegetation with a remote imaging system had to be solved in conjunction with the remote crop identification problem. If stresses, positive or negative deviations from the normal health of a crop, could be detected remotely, perhaps yield figures could be associated with the mapped stress and integrated over the entire field.

Two significant studies, by MacDonald(33) and Kumar and Silva(29), were generated from the NASA Corn Blight Experiment of 1971. This experiment employed both photographic cameras and a multispectral scanner to monitor the progress of the Southern Corn Blight in ground monitored fields during the summer of 1971*.

MacDonald(33) observed that stresses in corn that could be detected remotely included areas of low population (low leaf/ soil ratios), drought, hail damage, and weediness. Corn blight posed a detection problem because of its nature of spreading along the lower leaves of the stalk during its early stages.

---

*1970 was the year that Southern Corn Leaf Blight ravaged the United States corn crop which predominantly consisted of Texas male-sterile hybrid corn; destroying 15% of the total US corn produce.

34

Kumar and Silva(29) measured the emission and reflection of healthy and stressed plants by performing a computer analysis of data collected by an infrared spectroradiometer. The latter study noted a difference in the average emissivity-temperature (thermal infrared spectrum, 3-5u and 8-12u) between healthy and stressed plant populations. Supporting the ground based radiometry survey were the flights with the multispectral electro-optical scanner. Kumar reported that as the degree of stress increased, the stress was more statistically separable using combinations of one, two, three, or four spectral channels. The multispectral scanner band which best differentiated stressed from unstressed vegetation was the near-infrared band, 1-1.4u. Kumar hypothesized that this band might be optimal because the contrast between the soil reflection and corn leaf reflection was significant in this region of the spectrum. Kumar noted three bands of the multispectral scanner which best identified plant stress: 0.52-0.59u, 0.61-0.70u, and 1.0-1.4u.

Gausman(19) performed a laboratory photometry analysis of plant leaves and explained why stresses are detectable in the near-infrared portion of the spectrum. A stressed plant's near-infrared albedo is attributed to refractive index discontinuities between the hydrated cell walls and the intercellular air spaces and discontinuities among cellular constituents (i.e., membrane versus cytoplasm). Gausman noted that air

spaces of the leaves become larger as the plant matures. Simple optical physics accounts for an increased near-infrared reflectance as the air spaces of a leaf increase. However, to state whether a stressed plant has higher near-infrared reflectance or lower is dependent upon the stage of growth and chronological age of the plants whose spectra are being compared. Dehydrated diseased plants will have suffered tissue collapse and an increase in air voids; consequently, a greater near-infrared reflectance would be expected. Young plants which have been stunted in growth due to stress conditions, will have less mature leaf structure and smaller air spaces than neighboring mature plants. Consequently, a lower near-infrared spectral response would be expected. Gausman's work is supported by other subsidiary plant stress investigations (Ausmus(3)).

If yield prediction utilizing remote sensing technology were to be demonstrated, an optimal time for measuring the photosynthetic potential of a crop during a phase of the crop's growth cycle would have to be identified.

Phases of a growth cycle might be described as significant temporal divisions of a crop's maturation cycle during which some significant life process occurs. In the case of corn, Chirkov(10) divided the corn growth cycle into four phases:

36

(1) seedling to ten blades; (2) 11 blades to formation of
panicle; (3) formation of panicle to formation of corn fibers;
(4) flowering to milky ripeness. Similarly, the words germi-
nation, early development through pre-pollination, pollination,
and post-pollination development could have adequately des-
cribed the cycle. Nevertheless, remote sensing of the crop
during one or more of these phases may be important for
detecting stresses, identifying crops, and predicting the
yield of crops.

McKeever(36) has identified the critical period for
corn to occur at tasseling. The tasseling period is the
phase of the corn growth cycle when pollination occurs. If
hot dry weather is prevalent at pollination time, the shedding
of pollen, prior to the silk's emergence from the husks, is
hastened, causing a reduction in yield. This observation
is supported by Newlin(39) who stresses the importance of
humidity and temperature during the corn pollination phase.

The general conclusions from this aspect of the literature
review indicate the critical nature of weather at pollination
time. From personal communications with farmers, county
agricultural agents, and agronomists, the yield of corn is
not estimated until the weather conditions during pollination
are known. If weather or remotely acquired imagery data were

collected during a key phase like pollination, discounting

catastrophic events affecting crop yield (floods, fire, etc.)

perhaps crop yield estimates could be projected.

Steiner(49) identified the utility of the temporal

dimension for crop identification with an imagery data set

in 1969. Though his study was a photointerpretation study,

Steiner hypothesized that the time dimension would also be

important in digital image processing and automated crop

identification.   His study concluded that the interpretabil-

ity of land use from imagery was highly dependent on the

season and that no single season permitted identification

of crops with absolute certainty.*

The 1971 Corn Blight Experiment's results supported the

discovered importance of the temporal dimension for crop

identification purposes, MacDonald(33). Furthermore, studies

utilizing ERTS or LANDSAT satellite imagery have utilized

temporal registration of imagery for the purposes of demon-

strating improved automatic crop identification (Flores(16),

---

*The latter comment has been disproven at least once in the
case of opium poppy detection, Haynes(23). Haynes was able
to identify opium poppies, an important agricultural crop in
some countries, throughout the growth cycle, using only one
set of narrowband multispectral filters mounted on a multi-
spectral photographic sensor. The identification accuracy of
98% was demonstrated with imagery collected at both low and
high altitudes in three geographically and environmentally
distinct regions of the world. It is conceivable, however,
that other crops might not be as cooperative.

Carlson and Aspiazu(9)). Colwell(11) and Rachkulik's(43) investigations of soil-vegetation systems' albedo also supported the hypothesis that the time dimension might be quite valuable for estimation of the yield potential of a crop.

Synthesis of the Literature Review

Agricultural crops have dynamic spectral signatures which change during the growth cycle. Agricultural crops also have critical phases during their growth cycles, during which weather factors and other parameters critically mix to ultimately affect the yield of the crop. A measure of a crop's yield is its photosynthetic potential which can be determined from leaf area indexes, leaf to soil ratios, or multispectral brightness coefficients. The latter parameters can be detected from remote imaging systems. Finally, the literature emphasizes the importance of utilizing the temporal variation in imaged data for crop identification and yield prediction purposes.

Agricultural Crop Yield Prediction Program Overview

Bidirectional reflectance measurements of corn and other crops were acquired during the summer of 1974 using a tele-spectroradiometer developed by Spectral Data Corporation of Hauppauge, New York. The reflectance spectra were analyzed and subsequently, narrowband multispectral filters were selected for detecting levels of stress within a field of corn. Selected narrowband multispectral photography was acquired of eight agricultural test sites, as well as large areas of the Nittany Valley, Pennsylvania, during the 1974 corn-growing season. The purpose of the data collection was to provide multispectral photographic imagery of ground-monitored cornfields to the RADC Image Processing Facility for analysis in support of the yield prediction feasibility study. By both acquiring the unique imagery of periodically ground-monitored test sites throughout the corn growth cycle and by obtaining precise measurements of yield at the time of harvest, the feasibility of determining a relationship between the characteristics of a temporally-registered, multispectral data set and crop yield could be investigated.

The following procedures were used to acquire, process and analyze the selected narrowband multispectral photography:

40

1. The spectral reflectance measurements of healthy and hydrostressed corn had been acquired in August 1973 and subsequently, multispectral camera filters were selected for detecting stress within corn.

2. In 1974 additional spectral reflectance measurements of corn and other background crops were obtained and analyzed in order to establish that the bandpasses of the multispectral filters were correct for detecting stresses within corn.

3. A ground truth collection program was developed with the Pennsylvania State University Agronomy and Remote Sensing Departments to provide ground intelligence to verify corn-related phenomena observed on the aerial photographic imagery. This program also provided a measure of corn yield as a function of localized stressed areas within the target fields.

4. Aerial multispectral photography was collected over the test sites at low and high altitudes during April to October 1974. The resultant photographic negatives were processed under controlled conditions for analysis purposes.

5. Multispectral balanced positives, produced from the original multispectral negatives, were printed, processed and analyzed in an additive color viewer. Subsequent photo interpreter analyses of various additive color renditions were performed to establish the optimum color signatures for detecting stress in corn fields.

41

6. All of the multispectral imagery and related image and and ground truth were sent to the RADC Image Processing Facility. This film data was digitized, registered and stored on both magnetic disk and tape for further processing and analysis. Since the multispectral imagery was registered and spanned a time period of April to October, an extremely unique data set was available for demonstrating the feasibility of crop yield prediction.

7. Two parallel image processing studies were conducted at the RADC Image Processing Facility. The prime objective of the first study was to demonstrate the feasibility of predicting the yield of an agricultural crop, corn, by using a unique temporally-registered, narrowband multispectral data set. The second effort investigated the feasibility of semi-automatic crop identification using a narrowband multispectral data set that was not dedicated to speciating a particular crop; but rather for differentiating stresses within the crop.

8. To demonstrate yield prediction feasibility, an analysis of the time-variant multispectral data set of three test sites was conducted. For each test site's multispectral data set, areas corresponding to different types of stress, as reported from the ground truth program, were selected graphically with the computer. From the statistics inherent within the grey levels corresponding to the green, red, and infrared bands' imagery, a logic could be applied upon the

42

test set consisting of the entire corn field, so that a map of the stress patterns within the corn field could be computer generated.

9. From approximately 16 specific areas, corn minifields, within the digital **imagery of the three corn macrofield sites,** corresponding to geographic positions within the subject corn fields where accurate yield measurements had been made, the texture-tone feature statistics for the April-October imagery were extracted. After performing several multivariate and regression analyses of both this data and that collected during the ground truth program (i.e., weather, planting statistics, etc.), several important relationships pertinent to yield prediction were determined.

This entire study may be divided into two phases of scientific investigation. The first phase included the spectro-radiometry surveys and analysis, ground truth program, multi-spectral photography and supporting image truth collection programs, and the analogue analysis of the imagery utilizing additive color technology. The second phase incorporated the digital image processing effort, the stress detection study, the yield measurement study, the semiautomated crop identification study, and the yield prediction feasibility study. The yield prediction feasibility study included the multi-variate and regression analyses of the information extracted from the digitized photographic imagery.

43

A review of the results of the spectroradiometry survey, ground truth program, aerial photography collection program, and the additive color analyses is presented in the sections of this report which immediately follow. The techniques employed during both the digital image processing and the multivariate analyses of the data are presented in the main body of this report. Finally, a portion of this treatise presents both the review of the literature concerning crop identification studies and the results of an investigation, performed during this experiment, which investigated semi-automatic crop identification utilizing narrowband multi-spectral photography and digital image processing techniques.

Spectral Reflectance Characteristics of Corn

The technology of selected narrowband multispectral photography is based upon the experimentally derived fact that materials reflect light uniquely. In the case of living materials, agricultural crops, the spectral reflectance characteristics change as the crop species matures during the course of the growth cycle. Spectral reflectance differences exist between different agricultural crop species. Spectral differences also exist within an agricultural crop due to various positive and negative stresses. As defined in this report, agricultural stresses are positive or negative physiological deviations from the norm created by external causes. The analysis of an agricultural crop's spectral signature can establish the wavelengths and the magnitude of the reflectance differences both between and within a crop species. By selecting the narrowband multispectral filters at the appropriate wavelength bandpasses identified in the analysis of the crop spectral signature, the resultant multispectral photography, generated in conjunction with the filters, could be used to speciate the crop or detect stress patterns within the crop.

During a previous study in August 1973, the reflectance spectra of both normal and hydrostressed corn, alfalfa, and soybeans were acquired at an agricultural test site near Rome, New York (Grodewald,21). The objective of that particular

45

study was to demonstrate the feasibility of detecting hydro-
stress in corn. The objectives of this current study, during
which the spectroradiometry measurements were acquired
throughout the summer of 1974, were twofold: 1) To monitor
changes in the reflectance spectral signature of corn and
other crops during the growth cycle; 2) To re-evaluate the
selection of the bandpasses of the narrowband multispectral
filters utilized for the detection of stress within corn.

The significant observations and conclusions of this
spectroradiometry survey and data analysis are presented in
Appendix A, pages 391-408 of this report. The following section
describes the hardware of the telespectroradiometer utilized
for the acquisition of reflectance spectra of in situ agricul-
tural crops.

Instrumentation

The telespectroradiometer, Figure 23, was used to obtain
the in situ spectral signatures of normal corn, stressed corn,
and other agricultural crops. The three basic components of
the system are the detector head, control system and tele-
typewriter. A continuously variable interference filter
located in the detector head isolates the reflected radiation
at each wavelength. Two photomultipliers, for the visible and
near-infrared spectrums respectively, act as the detectors
of the Spectral Data Model 31 Telespectroradiometer. The
reflectance data is read into and stored in a minicomputer.
The output of the minicomputer is printed by a teletypewriter.

46
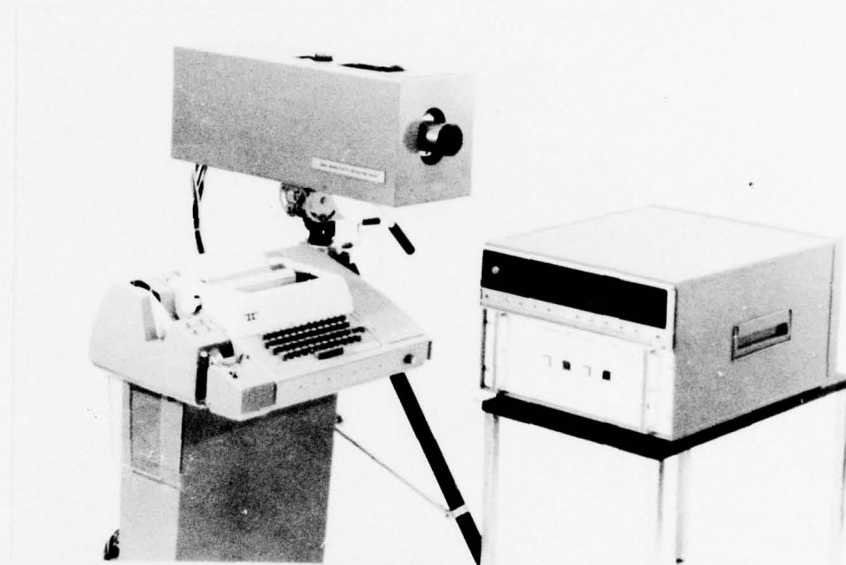
Figure 23.   Spectral Data Model 31 Telespectroradiometer used
             to obtain reflectance spectra of materials.

The telespectroradiometer was truck-mounted and driven to

the agricultural test sites located in Central Pennsylvania.

The boom-mounted radiometer permitted the acquisition of

reflectance spectra while inducing a minimal disturbance to

the natural environment. The detector head of the instrument

is located on top of a tower arm and is aimed at the target

material by remote control using a closed circuit television

system, Figure 24.

47
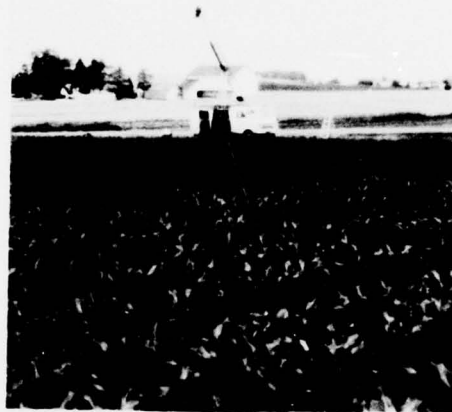
Figure 24. Truck-Mounted Telespectroradiometer.

Reflectance spectra of the target material and a standard
white reflectance plate were collected such that the tele-
spectroradiometer could be calibrated against the white ref-
lectance standard prior to sampling the spectra of the target
material.

The minicomputer's software computed the target material's
reflectance as a percentage of the solar radiation incident on
the white standard reflectance plate and removed the effects of
the photomultiplier dark current from the data. The percent
reflectance computations were printed on teletypewriter paper

48

for each target of interest. A single scan from 400 to 1075

nanometer(nm) was accomplished in twenty seconds. The band-

pass of the visible to near-infrared interference filters

was 12.5 nm. Above 1000 nm the sensitivity of the photo-

multiplier required filters with a bandwidth of 25 nm. The

field of view of the telespectroradiometer was normally two

degrees for all measurements.

Rationale for the
Multispectral Camera Filters' Bandpass Selection

When utilized as a data collection system for providing a data input to a crop yield prediction model based upon remotely-sensed imagery data, an operational sensor would be required to speciate crops rather than differentiate stresses within a crop. The reason for this necessary requirement is that it would not make sense to predict the yield of a crop that could not be identified. In support of this 1974 feasibility study in yield prediction, the results of the spectroradiometry survey were utilized to select four narrowband filters for detecting stresses within corn as opposed to speciating corn from other crops. Though the latter decision for filter selection may appear to contradict the previous statement concerning the nature of an operational crop identification-yield prediction sensor system, there is a rational explanation for the decision.

Two past studies performed by RADC related to agricultural crops. The first study in 1973 demonstrated the feasibility of accurate speciation of a particular crop, opium poppies, from other crops throughout the entire resolvable poppy growth cycle in three geographically distinct areas of the world(Haynes,23). The second investigation demonstrated the capability of the same imaging system, a narrowband multispectral camera using dif-ferent narrowband filters, to differentiate hydrostress from

normal healthy corn(Grodewald,21). A review of the liter-
ature concerning crop identification, presented on pages 340 -
353   of this report, indicated that accurate semiautomatic
crop identification has been demonstrated to only a mar-
ginally successful degree with respect to crop yield
prediction purposes, with the exception of the crop
speciation study which utilized the filter-specified, nar-
rowband, multispectral camera system(Haynes,23). Consequently,
it seems that remote crop identification may be a solvable
problem if the correct sensing tools are used. To develop an
experiment that would demonstrate the speciation of corn
during its entire growth cycle would be technically redundant
with respect to the opium poppy crop speciation program.
Instead, a greater challenge presented itself in the quest
to demonstrate the feasibility of crop yield prediction using
remotely-sensed data.

A preceding review of the literature concerning crop yield
prediction, pp: 30-39, indicated that the health and vigor of
a crop could be measured by the analysis of the amount of light
the crop reflected (Ref: 10,11,29,30,43,&45). The light
reflected from a crop varies as a function of the physiological
health of the crop or the degree of stress acting within the
crop. By integrating the technology expressed in the literature
relevant to crop identification and crop yield prediction, a

51

concept developed that a crop's yield might be predictable
by monitoring and analyzing the change in the crop's multi-
spectral albedo as a function of its maturation during the
growth cycle. Since it is well known that physiological
stresses directly affect the yield of a crop(Ref: 10,29&33),
it was determined that, for the purposes of the 1974 feasi-
bility study, it would be logical to dedicate the sen-
sor to detecting stresses within the crop rather than to nec-
essarily identify or speciate the crop. Once the stresses had
been detected and mapped by correlating and processing the
imagery with the ground truth data and image processing tech-
niques respectively, multivariate analysis of the image-
extracted data of known stressed areas within a crop would
identify any specific relationships between image data and
crop yield.

It is quite evident that if a relationship exists between
the change in reflectance of stressed areas within a crop of and
the yield of the same areas, then perhaps the analysis of the
light reflected in the narrow bandpasses utilized for crop
speciation could be used for yield prediction also. This
relationship would need to be determined in a later study.
In this initial investigation, the imagery data is dedicated
to detecting stress. Since stress is directly related to yield,
the multispectral imagery may have the best chance of being
related to a crop's yield through multivariate and regression
analysis.

Appendix A reviews the analysis and techniques used to determine the actual bandpass of the filters ultimately used for crop stress detection.

Multispectral Photography

Twelve multispectral photographic reconnaissance missions were flown by USAF RADC Flight Test aircraft over seven Pennsylvania State University agricultural test sites. Fifteen supporting image-truth-gathering photographic reconnaissance missions were flown in conjunction with the multispectral photography missions. These missions were accomplished between 29 April and 12 October 1974. During July and October of 1974, an AFFTC U-2 aircraft provided high altitude multispectral and conventional photographic coverage of the agricultural test sites. Weather, which generally varied from clear to partly cloudy with haze, was considered representative of that which would be encountered under normal operating conditions. Table 1 lists the relevant statistics of the twelve useable multispectral photography missions condensed from the total high and low altitude multispectral data collection program.

Multispectral Equipment

The Spectral Data Model 10 multispectral camera system, illustrated in Figure 25, was utilized in the collection of the photographic imagery. The camera body contains a four slit focal plane shutter assembly with a narrowband filter situated in front of each lens (150mm $\sim$ 6 inch focal length). In order

54

Mission No.	Date & Time Flown	Altitude (AGL)	Weather	Sites Photographed	Records	FILTERS Mean Wavelength (nm)	Bandpass (nm)
GR 74-14	29 Apr 74 1155-1317 EST	8,000 Ft	Clear	1,3,6,7	1 2 3 4	440 561 639 849	406-480 541-581 620-660 810-885
GR 74-17	1 May 74 1020-1310 EST	5,500 to 8,000 Ft.	Clear	1,3,4,6	1-4	Same As Above	Same As Above
GR 74-39	2 Jul 74 1014-1253 EST	6,000 Ft.	Clear	2,3,4,6, 7	"	"	"
GR 74-45	12 Jul 74 1010-1258 EST	25,000 to 50,000 Ft	Clear	1,2,3,4, 5,6,7	"	"	"
GR 74-47	16 Jul 74 1000-1220 EST	10,000 Ft	Cloudy	1,3,4,5, 6	"	"	"
GR 74-52	22 Jul74 1105-1235 EST	50,000 to 25,000 Ft	Clear	1,2,3,4, 5,6,7	"	"	"
GR 74-60	19 Aug 74 1127-1258 EST	6,000 Ft	Clear	1,2,3,4, 5,6,7	"	"	"

Continued on Next Page

TABLE 1: USEABLE MULTISPECTRAL PHOTOGRAPHIC RECONNAISSANCE MISSION PROFILES

55

Mission No.	Date & Time Flown	Altitude (AGL)	Weather	Sites Photographed	Records	FILTERS	
						Mean Wavelength (nm)	Bandpass (nm)
GR 74-65	4 Sep 74 1000-1110 EST	6,000 Ft	Cloudy	1,2,3,5,7	"	"	"
GR 74-69	15 Sep 74 1248-1334 EST	6,500 Ft	Cloudy	1,2,3,5,6,7	"	"	"
GR 74-79	4 Oct 74 1120-1335 EST	50,000 & 26,000 Ft	-	--	"	"	"
GR 74-83	8 Oct 74 1129-1314 EST	10,000 Ft	Light haze & clouds	1,4,6,7	"	"	"
GR 74-86	17 Oct 74 1153-1240 EST	8,300 Ft	Clear	1,2,3	"	"	"

TABLE 1 (Continued)

to achieve maximum image resolution, each camera lens can be
focussed for the wavelength bandpass transmitted by the
particular interference filter. The camera also has image
motion control and data annotation features.

Figure 25. Spectral Data Model 10 Multispectral Camera System

Based upon the analyses of both the reflectance spectra
acquired during the 1973 crop stress detection study(Grodewald,
21) and the reflectance spectra of corn stress obtained during
this 1974 program's spectroradiometry survey (Reference
Appendix A), a set of narrowband multispectral filters were
selected and updated to detect stress within corn throughout
the 1974 growing season. It should be mentioned that some
lag time occurred between the analysis of the reflectance data

and the actual utilization of the filters selected for the multispectral photography missions. This was primarily due to the fact that the photography missions were flown during the same growing season that the radiometry survey was performed.

The initial set of filters for the April 1974 multispectral photography missions were selected from an analysis of the reflectance data acquired during the 1973 experiment (Grodewald,21). It was a matter of good fortune that the filters originally selected from the 1973 data were determined to be adequate stress detecting filters for the entire 1974 corn growth cycle in Central Pennsylvania. The optical characteristics of these filters are tabularized in Table 2. The percent transmission of the filters as a function of wavelength is presented in Figure 26.

Record (Lens) No.	Peak Transmission (nm)	Half Peak Bandpass (nm)
1	440	406-480
2	561	541-581
3	639	620-660
4	849	810-885

Table 2. Optical Characteristics of the multispectral camera filters for corn stress detection.

58

Standard Kodak 2424 and special order Kodak SO-289
aerial black & white infrared films were used in the multi-
spectral camera. The multispectral photography was processed
at the Rome Air Development Center with a Kodak Versamat II-C
continuous film processor using MX 641 chemistry. Sensito-
metric control was maintained with an EG&G Mark IV sensitometer
for the purpose of matching the processing characteristics
of the film to the nominal film characteristics curve (density
v.s. exposure to white light), Figure 27.
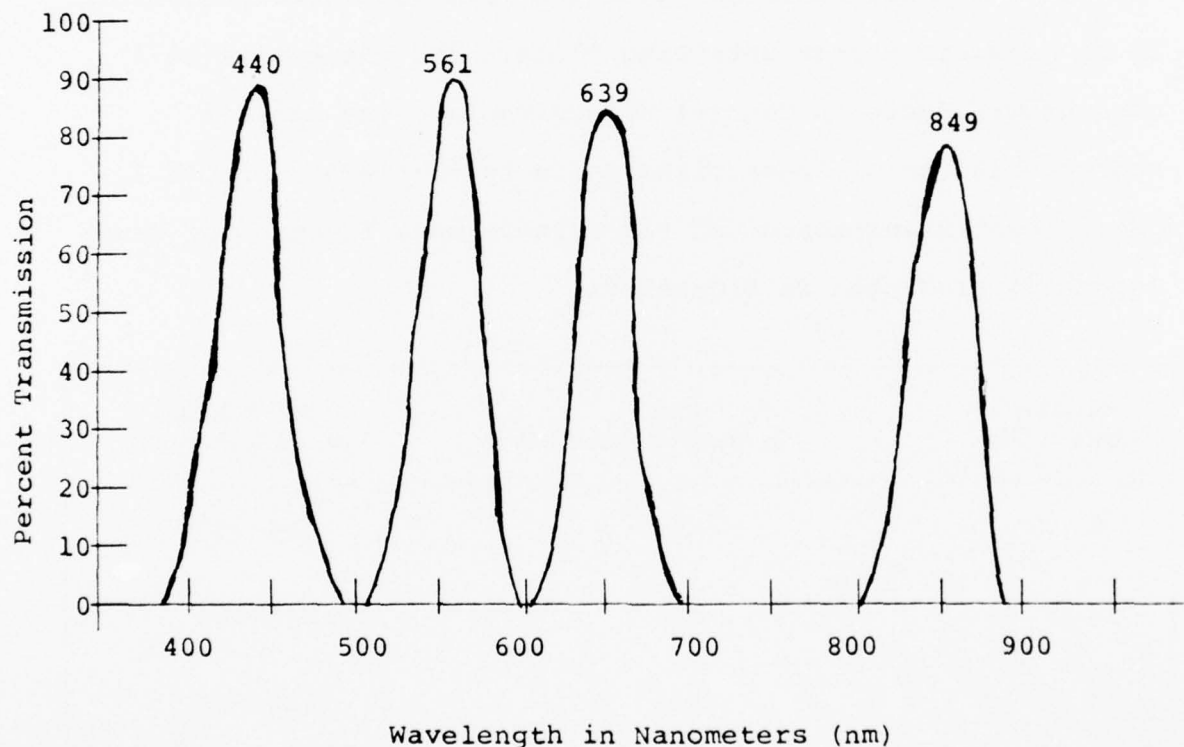
Figure 26. Percent Transmission of stress-detection multi-
spectral camera filters.

Positive transparencies having balanced exposure in all
four spectral bands' images were made with EK-2420 duplicating
film using a Spectral Data Model 41 optical projection printer.
This printer contains four lamps, the intensity of which can
be controlled independently, thereby allowing a considerable
amount of compensation for residual differential exposure errors
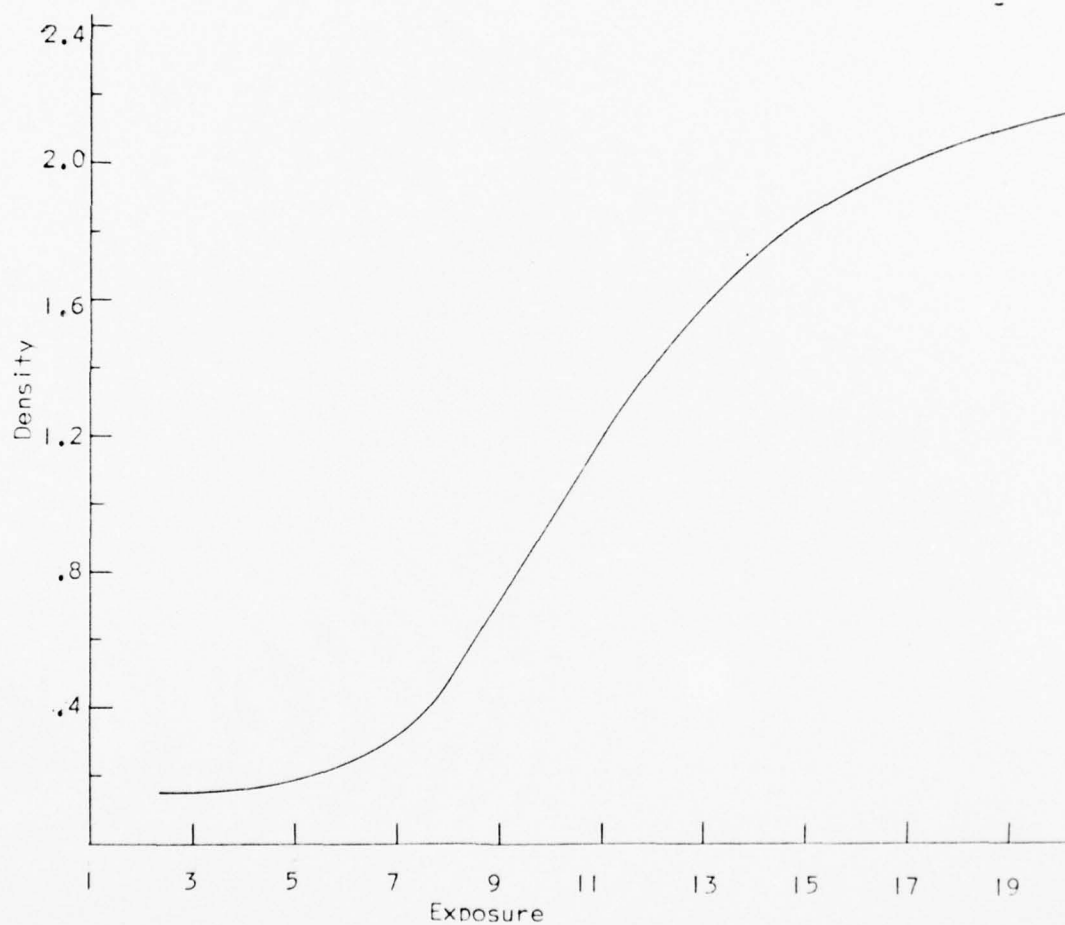
Figure 27.   Nominal processing curve for multispectral negatives.

which may exist between the original four multispectral negative images acquired by the airborne camera. Such differential exposure errors can result from changes in the conditions which existed when the aperture of the camera lens-filter combination was calibrated and those conditions which existed when the aerial multispectral photography was collected; i.e., distribution of the solar illumination, atmospheric haze, and changes in the reflectance of vegetation, soils, etc. Figure 28 illustrates the optical projection printer.
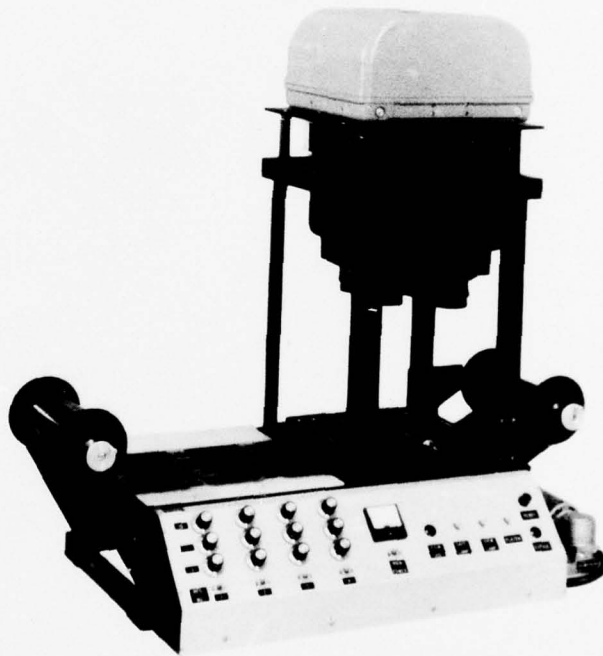
Figure 28.   Spectral Data Model 41 Multispectral Optical Printer.

The multispectral viewer, Figure 29, was used to recombine the balanced positives' black & white multispectral images using additive color technology(Wenderoth & Yost,54). This viewer contains four separate optical projection systems, one for each of the four black & white multispectral images. These projection optics each contain a long focal length scale correction lens in addition to the primary projection lens. The four illumination systems, one for each projection lens, contains sets of blue, green, and red filters. The imagery can also be viewed in black and white without filters. The resolution of the additive color viewer screen is such that a composite resolution of twenty-five lines per millimeter(mm) at four times enlargement of the photographic scale is produced. The additive color photographs presented in this report were produced by photographing the multispectral viewer screen with Ektacolor film.
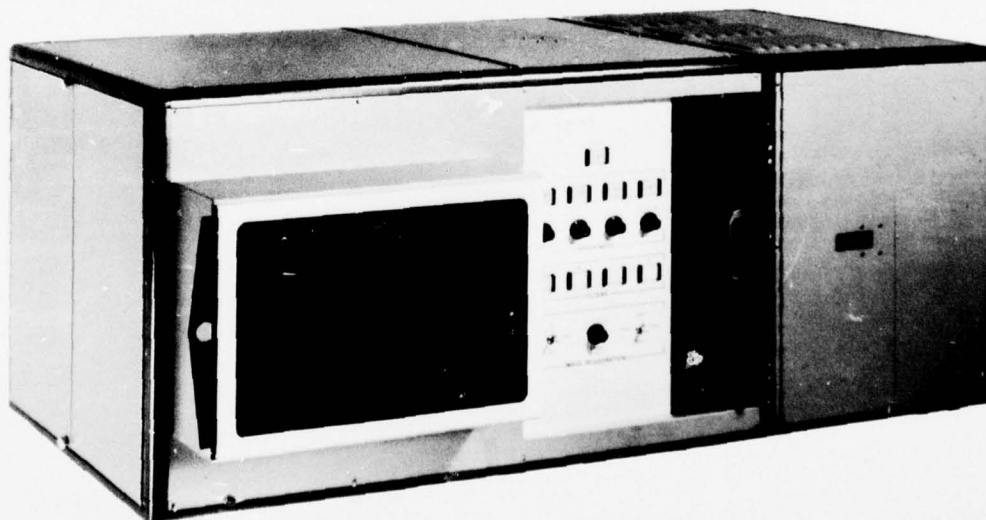
Figure 29.   Spectral Data Model 76 Additive Color Viewer.

A sample frame of black & white multispectral photography, typical of that collected over the Pennsylvania State University agricultural test sites, is presented in Figure 30. Though these photographs demonstrate that good image contrast was obtained in all four bands, the blue image typically had poor contrast and minimal information value.

The latter section which briefly described the multispectral sensor and analysis hardware was principally paraphrased from a section of an unpublished technical report (Grodewald,20), with the author's permission. The text entitled, "Multispectral Photography for Earth Resources", by Wenderoth and Yost(Ref: 54), is recommended for providing an excellent, in depth discussion of the technology of multispectral photography, additive color analysis, and the applications of such technology.

406-480 Nm Band

541-581 Nm Band

620-660 Nm Band

810-885 Nm Band

Figure 30: Multispectral photography collected over Field One
          at the Pennsylvania State University Agricultural
Farm, Pennsylvania. Taken at 1130 EDT on 2 July 1974. Altitude
6000 ft. AGL;. Scale 1:12,000 Mission: GR 74-039,Frame-61.

64

Ground Truth Collection Program

To support both the multispectral photography and image-truth collection efforts, a program for obtaining ground truth at seven large agricultural test sites was arranged with personnel of the Pennsylvania State University (PSU), Departments of Agronomy and Remote Sensing. A review of the ground truth collection program is presented along with the detailed ground truth of three of the seven agricultural test sites which were monitored and photographed during this feasibility study in crop yield prediction.

Figure 31 illustrates the geographic location of the seven agricultural test sites in Central Pennsylvania. These corn field test sites, with the exception of the 2.5 acre corn field (Site One) located at the PSU Agronomy Farm, varied in area from 40 to 150 acres. Sites One, Two, and Three were located within fifteen miles of the PSU Meteorology Department's weather station. Sites Four through Seven were located 40 to 100 miles from PSU and were weather-monitored by the Williamsport U.S. Weather Bureau Station. The selection of these fields was based chiefly on the requirement to provide agricultural sites with different soils and soil conditions.

The ground truth program was designed to provide detailed ground-monitored data collection on a microscale, mesoscale,
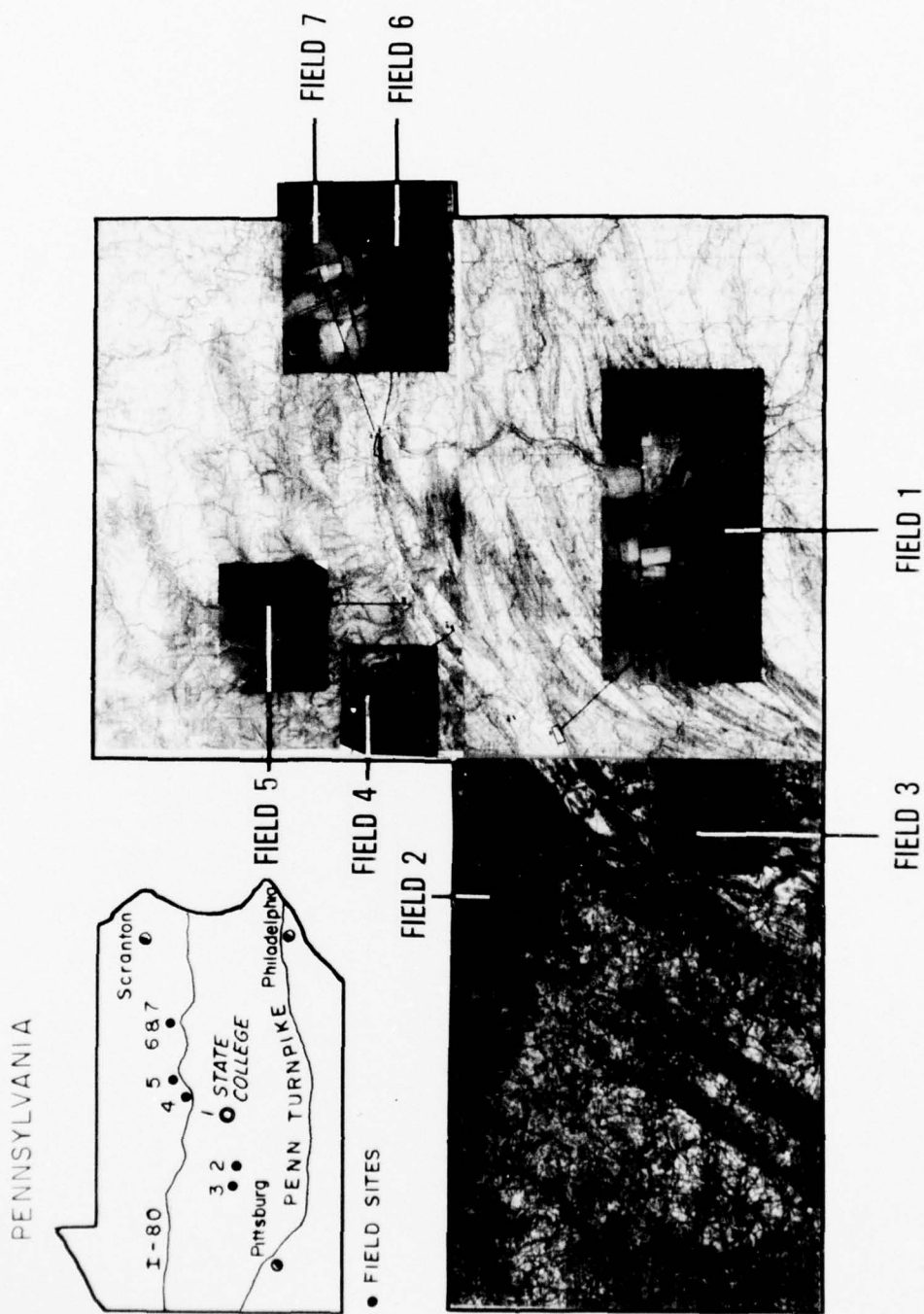
FIGURE 31 LOCATION OF PENNSYLVANIA STATE UNIVERSITY AGRICULTURAL TEST SITES

and macroscale. Microscale data was acquired from small plots at the PSU Agronomy Farm. Mesoscale ground truth was collected at the seven large agricultural test sites in Central Pennsylvania. Macroscale data generally included only yield data from selected farms in the Nittany Valley region of Pennsylvania which had been photographed at high altitude with both multispectral and conventional aerial cameras. Table 3 lists the different categories of information collected at the microscale and mesoscale levels of the ground truth program.

Weather Information

Air Temperature °F
Max Temperature, day °F
Min Temperature, day °F
Average Temperature,day °F
Dew Point °F
Wet Bulb °F
% Relative Humidity
    - 4 times daily
    - continuous*
Rainfall, inches
Visibility
Cloud Ceilings
Sky Cover
Wind Direction,daily*
Wind Speed,daily*
Illumination*
Evaporation*

Ground Truth Information

Crop Species/plot
Hybrid Information
Soil Test Results*
Soil Map
Soil Types
Soil Moisture Under sod*
Soil Temperature at dif-
    ferent levels below
    surface*
Base yield est./soil type
Row Spacing
Fertilization
Insect Control
Weed Control
Previous Crop
#Down Plants
#Barren Plants
Yield of crop/plot(bu/acre)
Stress Induction*
Stress Monitoring
--bimonthly progress
--cause of stress

Table 3: Microscale Ground Truth Information available at the PSU Agronomy Farm. Note that the asterisked(*) categories were not generally acquired for the mesoscale ground truth collection program which was preformed at the seven large agricultural test sites.

The PSU Agronomy Farm, located at Rock Springs, Pennsylvania, provided microscale, diurnal, detailed ground truth. Since many varieties of crops were grown in small plots at this farm, excellent spectral reflectance measurements of different crops were obtained here during this study. The many agricultural crops at the Agronomy Farm provided an ideal scenario for investigating the feasibility of semi-automatic crop speciation and stress detection using the multispectral camera system. The microscale ground truth also included the artificial induction of stresses in the corn planted in the small plots; i.e., Northern Corn Leaf Blight, hydrostress, etc. The effect of the stresses would be monitored through analysis of the multispectral photography. In summary, the PSU Agronomy Farm planted a variety of crops, permitted the acquisition of spectroradiometry measurements of the different crops during the entire growth cycle, provided detailed ground truth, and artificially-induced stresses into a number of small corn fields.

The mesoscale ground truth collection effort included seven large corn fields in Central Pennsylvania characterized by different soils, historic crop yields, planting conditions, and agricultural management practices. These fields were investigated on the ground every 1.5 weeks by a graduate research associate from the PSU Remote Sensing Department.

68

Besides acquiring the basic ground truth statistics at the beginning of the season, the student updated the records (i.e., fertilization, weed control, etc.) and monitored the stresses acting within each of the large corn fields. In order to accomplish the latter task of monitoring stresses, the student oriented himself in the corn field with annotated prints of black & white aerial photography of the site which were acquired during the image-truth collection program.

About a week before the student was to visit the sites, the aerial photography was collected over the test sites with either black & white infrared (Kodak 2424) or black & white (Kodak 2402) film. After the photography was processed and interpreted, the author annotated apparent stress areas within the corn field (apparent stress areas were selected subjectively by interpreting variations in film density within the imaged field). The annotated prints were sent to the graduate student to support his ground surveillance of the stresses within the corn field. During the site visits the graduate student would check the health of the corn in the annotated areas while referencing the annotated photography of the field. He would investigate the stressed areas imaged on the photography in detail so that the cause for the stress could be determined and recorded.

Besides the approximate biweekly monitoring of the large corn fields, the student also acquired yield estimates from

the owner of the farm. When considering the stressed areas

and the cause of the stress, the student would then make pro-

jections of the yield for that stressed area relative to the

farmer's estimate of the baseline yield for the entire field.

The farmer's estimates of the yield were recorded after the

weather conditions and field conditions at the pollination

stage of the corn growth cycle were known. This information

was utilized in selecting the regions in Sites One, Three, and

Seven where the yield was measured.

At the conclusion of the growth cycle, but prior to

harvest, the author selected specific areas within each of the

seven sites which had demonstrated a history of positive or

negative stress. These areas were annotated on black & white

photographic prints acquired from the most recent photographic

coverage of the test sites. During the first week of October

1974, the PSU student and a USAF lieutenant measured the yield

of one-five hundredth acre plots within the areas of stress

which had been annotated on the photography. The resultant

accurate measure of yield was used as the prime dependent

variable data which was later correlated with the multi-

spectral imagery data through multivariate analysis.

Finally, on a macroscale, imagery was collected at high

altitudes, 25,000 ft to 50,000 ft (AGL), over the Nittany

Valley of Central Pennsylvania. Besides the seven test sites,

the aerial coverage included fifteen other large farms which

had corn as the main crop. These extra sites provided no supporting ground truth, with the exception of the sites' yield for the 1974 corn growing season. These fifteen farms were members of the County Agricultural Extension Service's Corn Club. At the close of the season, members of the Corn Club reported their corn crop yields to the Extension Service along with planting data. Had more multispectral photography been acquired of these fifteen sites, the yield and image data would have been used to substantiate that the yield prediction model which developed from this study, does work when using high altitude data with no supporting ground truth. This data was not utilized as had been planned because an insufficient amount of multispectral photography was collected over the Nittany Valley Corn Club members during the 1974 season (U-2 coverage was collected in July and October 1974 only). However, the collected photography provided useable data for the subsidiary study in semiautomatic corn identification using multispectral imagery dedicated to corn stress detection (see Crop Identification section, pages 340-381)

Table 4 illustrates the multispectral film coverage collected over each of the seven agricultural test sites which were monitored during the mesoscale ground truth program. Despite the fact that adequate photographic coverage with the multispectral camera had been planned at the rate of one to two flights per week from April to October, only twelve missions'

71

fair to excellent photographic data was acquired. Fortunately, approximately fifteen good to excellent missions with conventional mapping cameras and various aerial films (2402-black & white, 2448-color, 2445-color, 2443-color infrared, and 2424-black & white infrared film) were flown during the same months thus providing image truth to support the multispectral data analysis. Pennsylvania was found to be plagued

Date of Mission	Mission #	AGRICULTURAL TEST SITES						
		1	2	3	4	5	6	7
29 Apr 74	GR 74-14	X		X			X	X
1 May 74	GR 74-17	X		X	X		X	
2 Jul 74	GR 74-39	X	X	X	X	X	X	X
12 Jul 74	GR 74-45	X	X	X	X	X	X	X
16 Jul 74	GR 74-47	X		X	X	X	X	
22 Jul 74	GR 74-52	X	X	X	X	X	X	X
19 Aug 74	GR 74-60	X	X	X	X	X	X	X
4 Sep 74	GR 74-65	X		X		X	X	X
16 Sep 74	GR 74-69			X		X	X	X
4 Oct 74	GR 74-79	X	X	X	X	X	X	X
8 Oct 74	GR 74-83	X			X		X	X
17 Oct 74	GR 74-86	X	X	X	X			
Total Number of Fair to Good MS Missions		11	7	11	9	8	11	9
U-Utilized Data R-Rejected Data		U	R	U	R	R	R	U

Table 4. Multispectral Photography Coverage for Seven Agricultural Test Sites in Central Pennsylvania.

by poor weather. Clouds, pilot error, camera failure, and camera operator errors were some of the problems that prevented the data collection effort from being a complete success.

Table 4 illustrates the fact that of the seven test sites, only Sites One, Three, Four, Six, and Seven had adequate multispectral photographic coverage during the entire growth cycle (nine or more good to excellent missions). However, Site Four's imagery had to be rejected due to scattered clouds eliminating too much of the useable area of the field in several "useable" frames of photography. Field Six was a huge site having excellent ground truth data and numerous areas where yield had been sampled. However, it too was rejected for having inadequate "useable" photographic imagery due to both weather conditions and pilot error. Consequently, only Sites One, Three, and Seven were chosen for the image processing and multivariate analysis study in yield prediction. These three sites together had a total of sixteen areas where yield had been measured and both supporting image and ground truth had been collected. The detailed ground truth of Fields One, Three, and Seven is presented immediately following this section. A rough sketch of the ground truth of the other four mesoscale corn field sites is presented in Appendix B of this report. The detailed ground truth and the photography resulting from the additive color analysis of the fields is presented in an unpublished draft report printed by RADC(Grodewald,20).
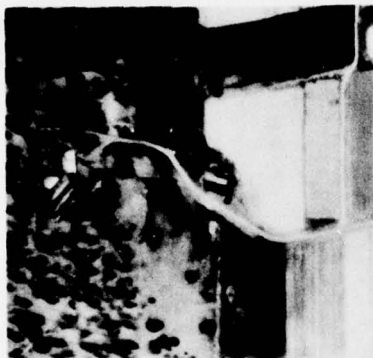
73

1 May 1974 5500 Ft AGL   2 July 1974 6000 Ft AGL

16 July 1974 10000 Ft AGL   19 August 1974 10000 Ft AGL

4 September 1974 6000 Ft AGL   8 October 1974 10000 Ft AGL

Figure 32: Six sequential multispectral additive color images of Site One; Multispectral black & white photography collected on the respective dates generated the additive color renditions on the SDC additive color viewer when the green image was color projected as green, the red image as red, and the infrared image as blue. Note the stress patterns within the field which are most noticeable in the 2 July, 16 July and 4 September images.

74

Figure 33. Field One in June 1974.

Figure 34. Typical healthy corn at the Pennsylvania State University Agronomy Farm in July 1974.

<u>Detailed Ground Truth</u>: Site One

Site One, Figures 32, 33 & 35, is located at the Pennsyl-
vania State University Agricultural Farm at Rocksprings,
Pennsylvania. More multispectral photography was taken over
this field than any other field during the project.

<u>Field One Soil Data</u>
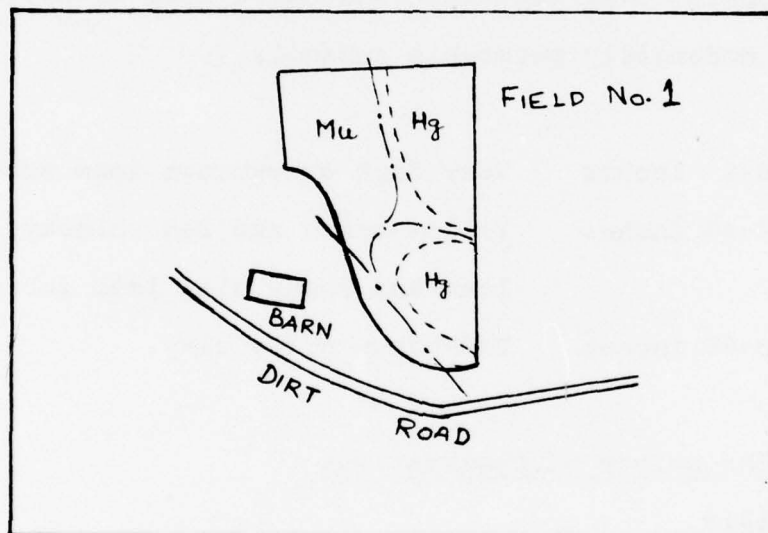
Figure 35. Soil map of Field One.

<u>Hagerstown Silt Loam (Hg)</u>: A deep well drained soil on uplands,
having moderately permeable subsoils, formed from weathered
limestone.

Profile: 0-14 inches   Dark to red-brown granular silt loam
                         surface soil.

         14-85 inches   Yellow-red and brown, blocky, silty
                         clay subsoil.

           85 inches   Limestone bedrock.

Merrill Silt Loam (Mu): Deep, well-drained soil on uplands
formed, in materials weathered from colluvium containing sand-
stone and shale, over materials weathered from limestone. These
soils have moderately permeable subsoils.

Profile: 0-8 inches   Very dark grey-brown loam surface soil.
          8-40 inches   Yellow-brown and red, blocky, gravelly
                         loam and sandy clay loam subsoil.
         40-66 inches   Red-brown silty clay.

Field One Chronology of Observations

   1 May 1974

   Field One planted with Funks 4445 variety of corn, row
spacing 38 inches, plant population of field 26,500. No manure
or lime had been applied, however, 500 pounds per acre of 0-15-30
(% N-P-K) and 150 pounds of anhydrous (or aquevin) nitrogen were
both plowed into the field in the spring. In addition, 100
pounds per acre of 15-40-5 (% N-P-K) was deposited by the

planter along with seed. Pre-emergence weed control was accom-
plished using 2# atrazine and 1.5# simazine; producing good
results. Insect control was accomplished using .75# Furidan
at planting time.

8 July 1974

The field has two primary soil factors influencing the
growth of the corn. They are rockiness or shallow depth to
bedrock and surface textural changes. The soil is a limestone
residual soil with areas where bedrock is near the surface
(due to non-uniform weathering of the limestone). The sur-
face textural changes are important in those areas where
clay content is greater. Both conditions limit the soil's
moisture holding capacity and/or availability of that moisture
to plants.

Corn height as of 8 July 1974:

Location	Height
1	37-43"
2	55-62"
3	50-56"
4	60-65"

Field No. 1
8 July 1974
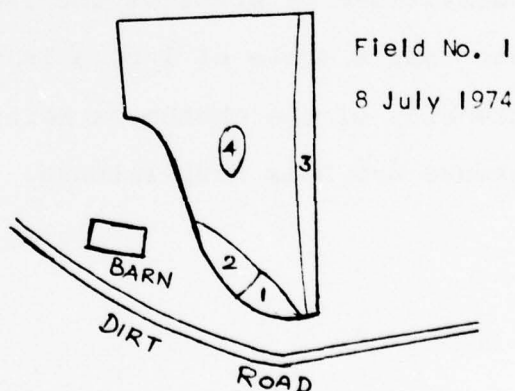
23 July 1974

Corn doing well; no evidence of disease or insect damage; corn height varies from 6 to 8 feet.

8 August 1974

Corn doing well; no insect or disease damage evident; corn height variable from 6 to 8 feet.

22 August 1974

Corn doing well with variable height from 6 to 8 feet; some northern corn leaf blight (or Stewart's) disease evidenced by small lesions on leaves.

9 September 1974

Estimated yield is between 115 and 125 bushels per acre. Observations on areas of the field are listed below by area. Note that a scale of 1 to 5 is used to indicate relative intensity of the phenomena being observed where 1 is least intense and 5 is most intense.
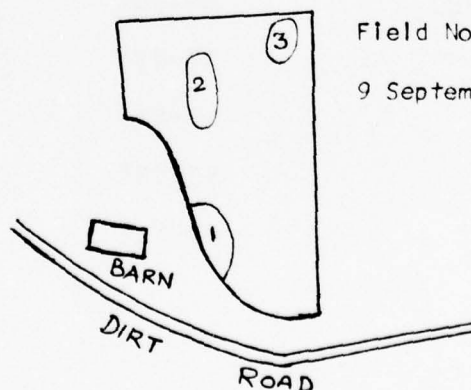
Field No. 1

9 September 1974

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
1.	Nutrient Problem	None	<1	As Planted	7-8	105-115

Weeds <1, corn has some brown leaves (≤ 1-2).

2.	Topo. depression	None	<1	As planted	6-7	90-100

Weeds <1, corn very green, some deer damage (1.5-2.5).

3.	Eroded knoll	None	<1	As planted	7-8	110-120

Weeds none; corn very green and vigorous.

23 September 1974

Estimated yield is between 115 and 125 bushels per acre.
Observations on areas of the field are listed on the following
page by area number, intensity of phenomena observed on a
scale of 1 (least) to 5 (most).

Field No. 1

23 September 1974

80
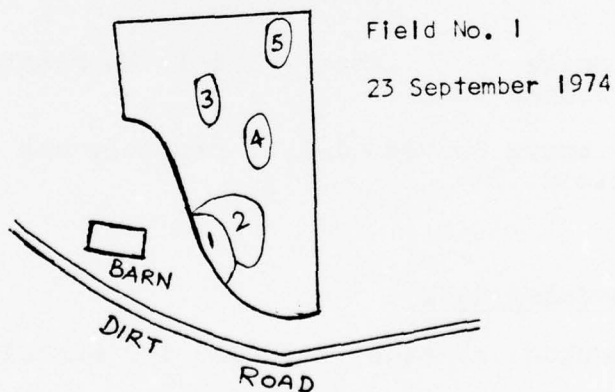
Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
1	Nutrient problem	None	∠1	As planted	6.5-7.5	70-80

Weeds 0; many brown leaves 3.5-4.5; probably the poorest area of the field.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
2	Better moisture	None	∠1	As planted	7.0-8.0	90-100

Weeds 0-1; few brown leaves <1-2.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
3	Topo. depression	None	<1-1	As planted	6.5-7.5	75-85

Weeds 0; brown leaves 1-2; some deer damage.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
4	Soils slight textural change (more clay)	None	∠1	As planted	6.5-7.5	80-90

Weeds 0-1; few brown leaves <1; some plants fallen over perhaps due to deer.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
5	Soils eroded knoll	None	<1-1	As planted	6.5-7.5	90-100

Weeds 0; brown leaves <1-1.5; probably one of the best areas of the field.

2 November 1974

Harvested; average yield 129 1/2 bushels per acre; range 107.6 to 157.0 bushels per acre.

81

Field One Yield Data
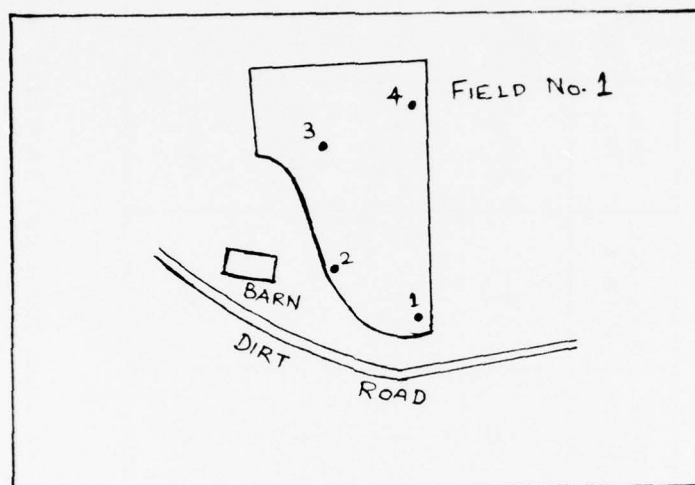
Figure 36. Locations in Field One of the Pennsylvania State
University Agronomy Farm where yield measurements were
collected.

Harvest Date: 2 November 1974

Average Yield: 129.5 bushels per acre

Crops: Corn, Funks 4445 variety

Length of Row Sampled: 27 feet 6 inches

Row Spacing: 38 inches

Sample Area: 1/500 acre

82

Loc.	Gross Wt.	Tare	Net Wt.	# Plants	# Down Plants	# Barren Plants	Wet Wt.	Dry Wt.	Yield Bu/A
1	40 3/4	14	26 3/4	49	0	3	940.6	538.1	129.4
2	33 1/4	14	19 1/4	42	2	3	638.0	421.8	107.6
3	38 3/4	14	24 3/4	55	3	3	810.8	480.1	123.9
4	42 3/4	14	28 3/4	61	2	4	714.4	461.4	157.0

Table 5. Yield measurements at four locations in Field One.

83
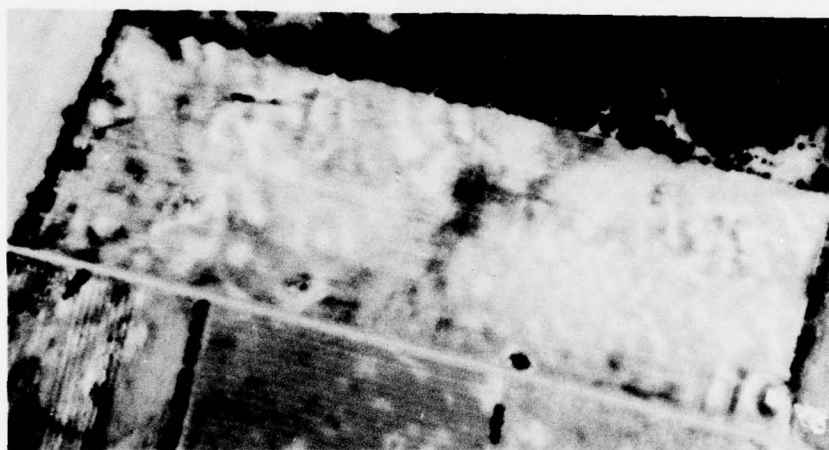
Figure 37. Multispectral additive-color rendition of Field
Site 3. Multispectral black & white photography collected
on 4 September 1974 from an altitude of 6000 ft (AGL)
generates the additive color rendition on the SDC additive
color viewer when the green image is color projected as
green, the red image as red, and the infrared image as
blue. Notice the stress patterns within the field that
are apprently enhanced.

<u>Detailed Ground Truth</u>: Site Three

Site Three, Figures 37 & 38, is located approximately 18 miles southwest of the State College, Pennsylvania, near the town of Spruce Creek. This field had previously been planted in hay. Corn growth was rather poor throughout the season, the yield being less than 90 bushels per acre.

<u>Field Three Soil Data</u>
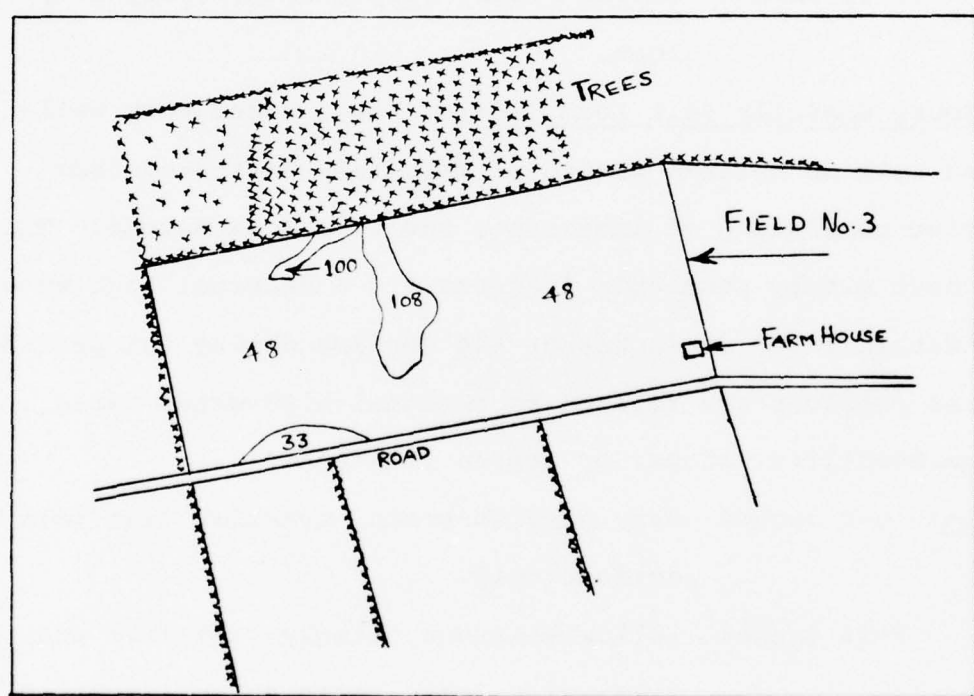
Figure 38. Soil map of Field Three.

Hublersburg Cherty Silt Loam (48): Deep, well-drained soils on uplands formed in materials weathered from siliceous limestone and dolomite. These soils have moderately permeable subsoils. Most use problems are related to coarse fragments or slope.

Profile: 0-8 inches Dark grayish-brown to brown, granular, cherty silt loam surface soil.

8-51 inches Strong brown, blocky, silty clay loam subsoil.

51-61 inches Strong brown, very cherty, silty clay loam.

Clarksburg Gravelly Silt Loam (108): Deep, moderately well-drained soil on uplands formed in materials weathered from colluvium consisting of limestone, shale, and sandstone. These soils have slowly permeable fragipan and a seasonal high water table within 18 to 36 inches of the surface during wet periods. Most use problems are related to seasonal high water table, slow permeability, slope, or coarse fragments.

Profile: 0-9 inches Dark grayish-brown, granular silt loam surface soil.

9-28 inches Yellowish-brown, blocky silt loam and silty clay loam, upper subsoil mottled grayish brown below 23 inches.

28-48 inches Yellowish-brown, prismatic silty clay loam, lower subsoil mottled gray and strong brown.

86

Hagerstown Silt Loam (33): A deep well-drained soil on uplands formed in materials weathered from limestone having moderately permeable subsoils.

Profile: 0-14 inches   Dark to red-brown granular silt loam surface soil.

14-85 inches   Yellow-red and brown, blocky, silty clay subsoil.

85 inches   Limestone bedrock.

Nolin Silt Loam (100): Deep, well-drained soils on flood-plains formed in stream deposits washed from uplands, under-lain by limestone, sandstone, siltstone, and shale. These soils have moderately permeable subsoils. Most use problems are related to flooding hazard.

Profile: 0-8 inches   Dark grayish-brown, granular silt loam surface soil.

8-52 inches   Dark yellowish-brown and brown, blocky silt loam subsoil.

Field Three: Chronology of Observations

27 April 1974

Field Three was planted with Pioneer 3773 variety of corn, row spacing - 34 inches, plant population - 23,000. No manure or lime applied; 104 pounds of nitrogen, 52 pounds of potassium, and 52 pounds of phosphorous in dry form applied during spring planting; 300 pounds of 15-10-10 (% N-P-K) applied during

planting. The previous crop was hay (and weeds). Pre-emergence
weed control was 1.25 pounds of Attrex and 1.5 quarts of Lasso;
post-emergence weed control using 1.5 pounds of Attrex. Ten
pounds of Dyfonate 10G was used for insect control.

8 July 1974
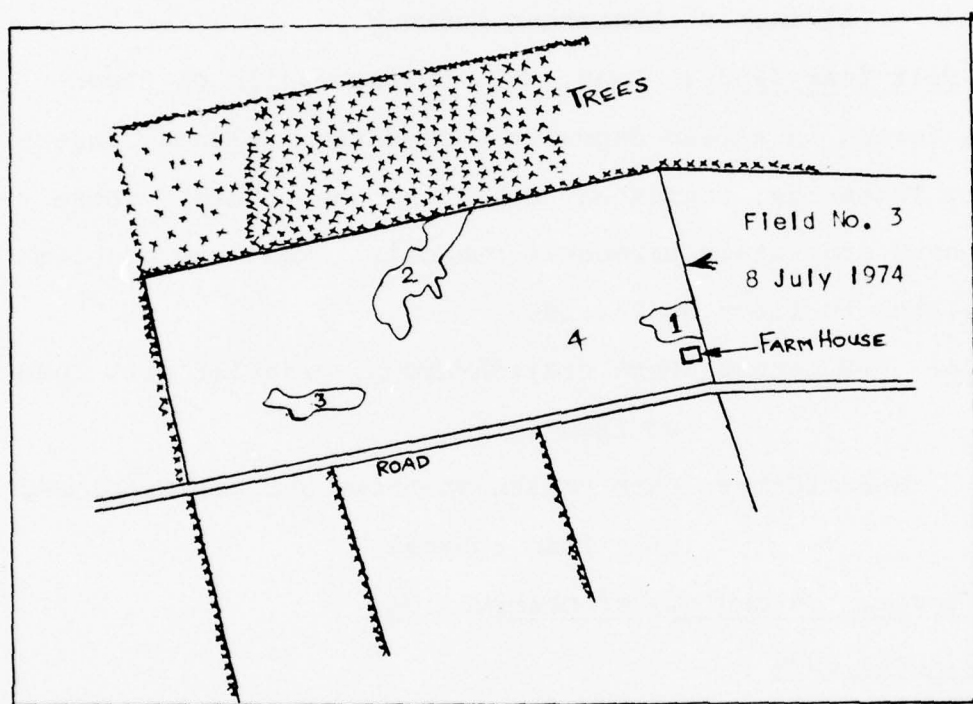
Location

Observations

1.   This area has a soil which is shallow to bedrock
(approx. 12"). Consequently, the soil has a low
moisture-holding capacity. Corn is doing very
poorly and is a pale green color.

Location	Observations
2.	This is topographically the lowest area of the field. The soil is a deep, moderately well-drained Clarksburg silt loam. Therefore, it generally has a higher moisture content and, during periods of low-to-moderate rainfall, would allow the corn to do fairly well. However, it is slowly permeable and has a seasonal high water table and during periods of high rainfall, the corn would suffer. Corn height is about 8 feet.
3.	Corn is good in this area. Soil is Hublersburg cherty silt loam. Corn height is about 4.5 feet.
4.	Corn is a variable height in this area. Depth to bedrock is variable and therefore corn height is extremely variable.

General Comments: The soils in this area are limestone residual soils and have highly variable depths to bedrock. Where the soil is shallow, the corn does poorly and where it is deep, the corn does well.

23 July 1974

There was little change in the corn relative to 8 July 1974.

8 August 1974

A more detailed investigation of those areas where the corn appears stunted revealed that other factors seem to be causing the differences. Previously, the shallow depth to bedrock was suspected of influencing the corn. However, the soil depth does not seem to be the cause in all cases. The corn is displaying what appears to be a magnesium deficiency.

Overall, the corn displayed little change with no disease or insect damage.

22 August 1974

The overall condition of the corn has not changed and a series of soil samples were taken for laboratory analysis. The area marked #2 on the sketch below is an area where the corn is doing very well with corn height in excess of 8 feet. No disease damage evident.

9 September 1974

Estimated yield is between 85 and 95 bushels per acre. Observations on portions of the field are listed below by area. A scale of 1 (least) to 5 (most) is used to indicate the relative intensity of the phenomena being observed.
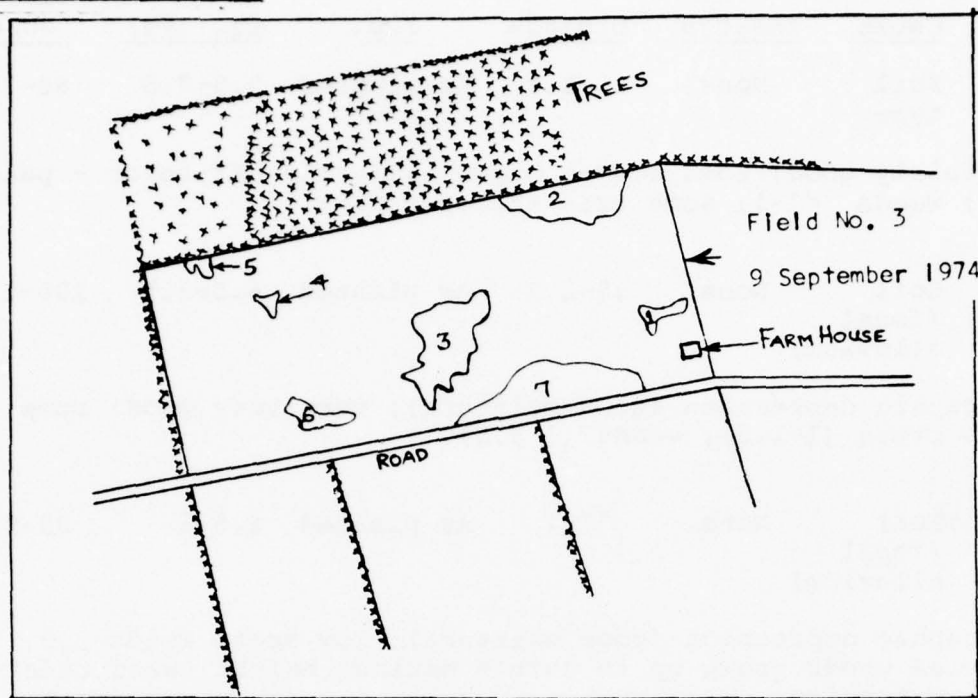
Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
1	Soil eroded	None	.5-1.5	As planted	4-6	15-30

Soil has higher coarse fragment content, corn extremely poor; very small ears; many leaves brown (2-3); possibly some frost damage; high soil response; weeds < 1.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
2	Soil type	None	.5-1.5	As planted	5.5-7.5	60-75

Corn fairly good; some leaves brown; slightly off-color - pale green; weeds <1-1; some tree shadow effect.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
3	Soil (local alluvium)	None	.5-1.5	As planted	6.5-8.5	100-110

Topographic depression (good moisture); corn very good; some leaves brown (1-1.5); weeds (1-2.5).

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
4	Soil (local alluvium)	None	<1-1	As planted	6.5-8	85-95+

Topographic depression (good moisture); low spot; weeds 3.5; some vine weeds grown up to corn's maximum height; weed undergrowth extensive.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
5	Eroded knoll	None	.5-1.5	As planted	4.5-7	30-45

Soil has much higher coarse fragment content; exposed B horizon material; corn slightly pale green; weeds (1-2.5).

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
6	Good soil on side slope	None	.5-1.5	As planted	6.5-7.5	95-105

Corn better than average for field; weeds <1; brown leaves 1-2.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
7	Good soil	None	<1	As planted	6.5-9	110-120

Corn very good; leaves brown (2-2.5); better than field average.

23 September 1974

Field No. 3

23 September 1974

FARM HOUSE

TREES

ROAD
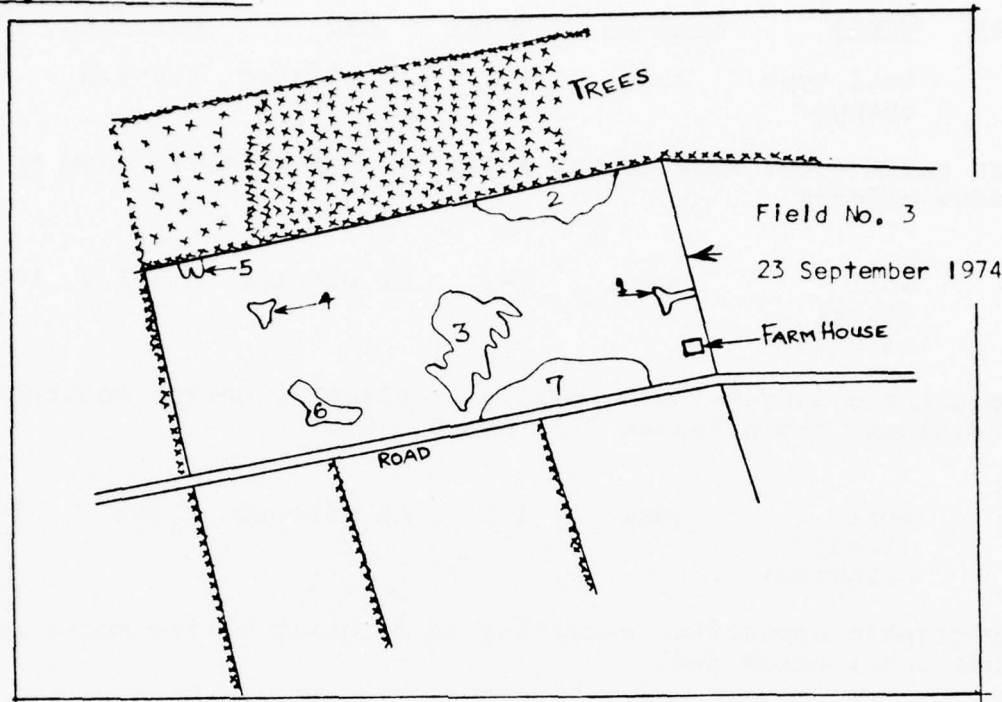
23 September 1974

Estimated yield is 90 to 105 bushels per acre. Observations
on areas of the field are marked below by number; intensity of
phenomena observed on a scale of 1 (least) to 5 (most).

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
1	Eroded area and/or relatively shallow	None	1-2	As planted	4-6	<15-30

Higher coarse fragment content in soil; corn extremely poor;
corn ears very small; almost totally brown leaves 4.5-5; very
high soil response on photos; weeds <1.

93

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
2	Soil type change	None	1-2	As planted	5.5-7.5	60-75

Corn quite good; some leaves brown 1-2; weeds <1-1; some tree shadow effects.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
3	Soil (local alluvium)	None	1-2	As planted	6.5-8.5	100-110

Topographic depression resulting in slightly better moisture conditions; brown leaves 2-3; weeds 1-2.5.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
4	Soil (local alluvium)	None	1-2	As planted	6.5-8	90-100

Topographic depression resulting in slightly better moisture conditions; weeds 3-5.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
5	Eroded knoll (exposed B horizon)	None	1-2	As planted	4.5-7	40-50

Much higher coarse fragment content (lower moisture holding capacity); brown leaves <1-2; weeds 1-2.5.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
6	Good soil (slight sideslope)	None	1-2	As planted	6.5-7.5	95-105

Some local alluvium in soil; corn better than average for field; weeds <1; brown leaves 2-3.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
7	Good soil	None	1-2	As planted	6.5-9	110-120

Corn very good (probably best area of field); brown leaves 3-4.

2 November 1974

Field harvested; average yield 89.7 bushels per acre. Range 63.4 to 120.0 bushels per acre.

94

Field Three Yield Data
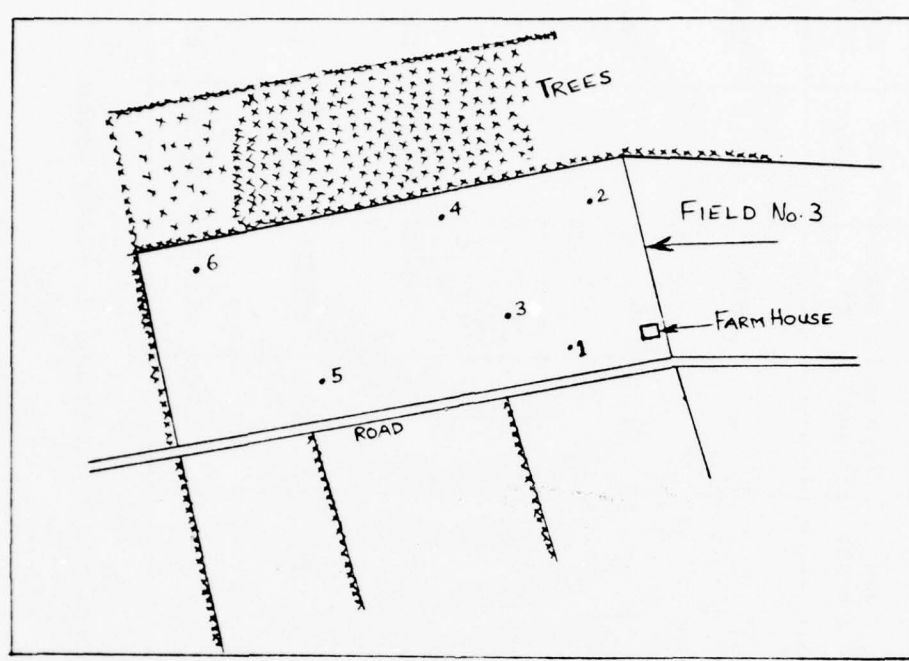
Figure 39. Locations in Field Three where yield measurements
were collected.

Harvest Date: 2 November 1974

Average Yield: 89.7 bushels per acre

Crops: Corn, Pioneer 3773 variety

Length of Rows Sampled: 30 feet 10 inches

Row Spacing: 34 inches

Sample Area: 1/500 acre

Loc.	Gross Wt.	Tare	Net Wt.	# Plants	Down Plants	Barren Plants	Wet Wt.	Dry Wt.	Yield BU/A
1	33 1/4	14	19 1/4	39	0	1	1079.3	606.9	91.5
2	31 1/4	14	17 1/4	42	0	4	738.8	397.2	78.4
3	14 1/2	14	1/2	37	0	33	-	-	-
4	39 1/4	14	25 1/4	41	0	0	1066.8	600.0	120.0
5	33 3/4	14	19 3/4	36	1	3	1043.7	594.8	95.1
6	29 3/4	14	15 3/4	41	1	5	654.2	311.3	63.4

Table 6. Yield measurements at six locations in Field Three.

96
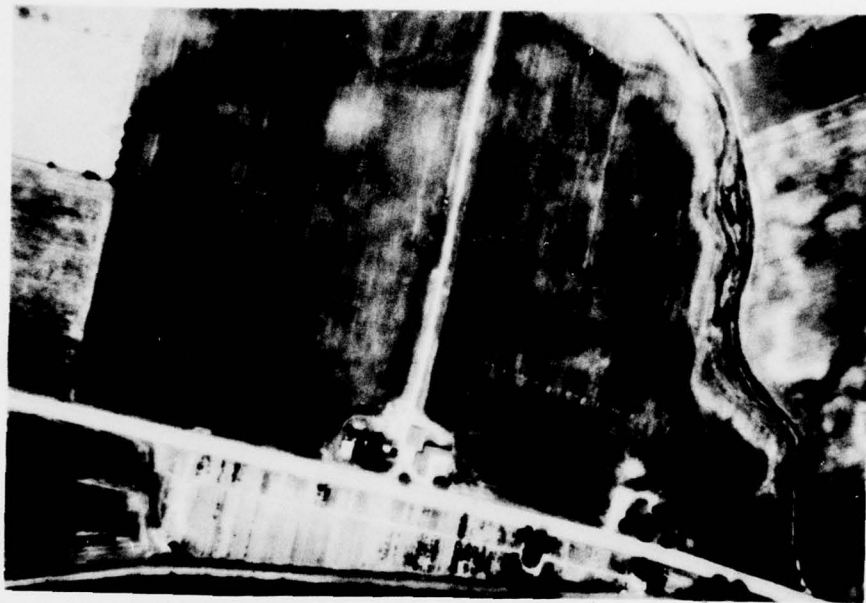
Figure 40. Multispectral additive-color rendition of Field
Site 7. Multispectral black & white photography collected on
2 July 1974 from an altitude of 6000 ft (AGL) generates the
additive color rendition on the SDC additive color viewer
when the green image is color projected as green, the red
image as red, and the infrared image as blue.

Detailed Ground Truth: Site Seven

Site Seven, Figures 40 & 41, is located adjacent to Field Six, fifteen miles east of Williamsport, Pennsylvania, near the town of Muncy. The field is two miles east of the west branch of the Susquehanna River.

Field Seven Soil Data
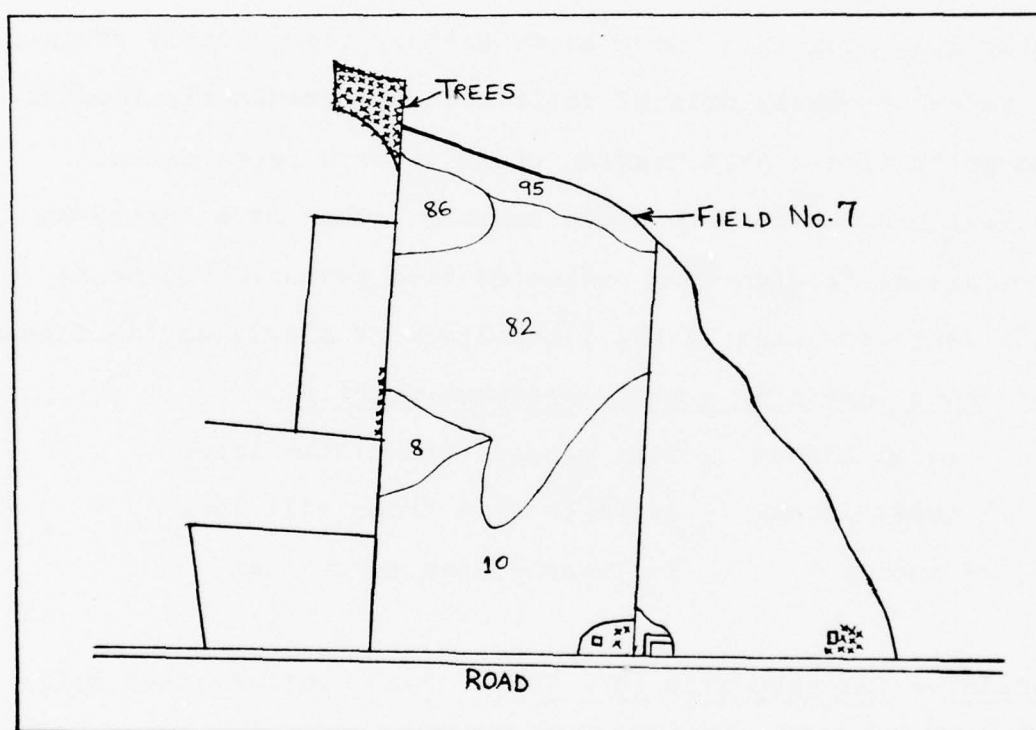
Figure 41. Soil Map of Field Seven.

98

Allenwood Gravelly Silt Loam (82): Deep, well-drained soils
that developed from glacial till. The till is composed of gray
sandstone, shale, and quartzite.

Major layers in a representative profile:

0 to 15 inches	Yellowish-brown gravelly silt loam
15 to 42 inches	Yellowish-red gravelly, silty, clay loam
42 inches	Very gravelly silt loam

Holly Silt Loam (8): Deep to moderately deep, poorly drained
to somewhat poorly drained soils that are frequently flooded.
The soils have a dark grayish-brown surface layer and a
yellowish-brown to pale-brown subsoil. They have formed on
floodplains from uniform medium to fine textured sediments
that were deposited by the floodwaters of slowly moving streams.

Major layers in a representative profile:

0 to 13 inches	Dark grayish-brown silt loam
13 to 24 inches	Brown to pale brown silt loam
24 inches	Yellowish-brown sandy loam

Hartleton Channery Silt Loam (86): Deep, well-drained soils
formed mainly in material derived from acid, gray, and brown
shale. In places, the material was thoroughly mixed by
glaciation with smaller amounts of sandstone and darker shale.

99

Major layers in a representative profile:

0 to 7 inches   Grayish-brown to yellowish-brown channery silt loam

7 to 36 inches   Yellowish-brown channery silt loam to dark brown channery loam

36 to 45 inches   Coarse shale fragments

Watson Silt Loam (95): Deep, moderately well-drained soils developed in glacial till or in periglacial and colluvial deposits in a mixture of material from gray shale, siltstone, and sandstone. Gently sloping areas are occupied by seep water from higher slopes during wet periods.

Major layers in a representative profile:

0 to 10 inches   Dark yellowish-brown silt loam

10 to 29 inches   Reddish-brown silty clay loam to reddish-brown clay loam

29 to 36 inches   Strong brown clay loam

Philo Silt Loam (10): Deep, moderately well-drained soils on floodplains formed in stream deposits washed from uplands underlain by gray sandstone and shale. These soils have moderately to moderately-slow permeable subsoils and a seasonal high water table within 18 and 36 inches of the surface during wet periods. Most use problems are related to seasonal high water table, flooding hazard or moderately slow permeability.

100

<u>Major layers in a representative profile</u>:

0 to 10 inches   Very dark, grayish-brown silt loam
                  surface soil

10 to 39 inches   Dark brown to dark yellowish-brown blocky
                  silt loam subsoil mottled gray below 19"

39 to 52 inches   Stratified sand, silt and gravel

<u>Field Seven, Chronology of Observations</u>

<u>7 May 1974</u>

Field Seven planted with Pioneer 3369A, 3517 and 3518
varieties of corn, row spacing 38 inches, plant population
20,000. No manure and no lime applied. Regular 10-10-10
(% N-P-K) applied, 300 pounds per acre, during spring planting
and 300 pounds of 10-10-10 applied during planting. Previous
crop was corn. An unspecified type of post-emergence weed
control was used. Insect control was accomplished using
Dyfonate.

7 May 1974
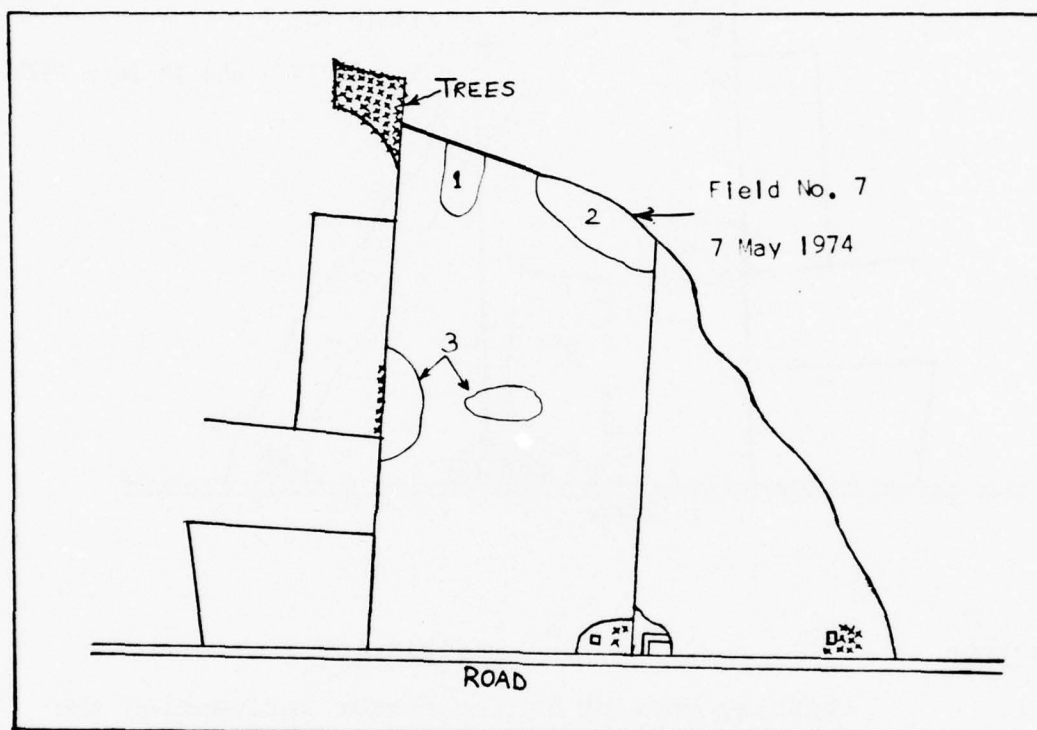
Area 1 contains a winter spring which appears to be not too
severe. Area 2 contains heavier soil (more clay). Area 3
is a recess in the field which is about 75 feet long, 40 feet
wide, and has a maximum depth of about 10 feet. It is of
unknown cause.

9 July 1974

TREES

Field No. 7

9 July 1974 and 24 July 1974

ROAD
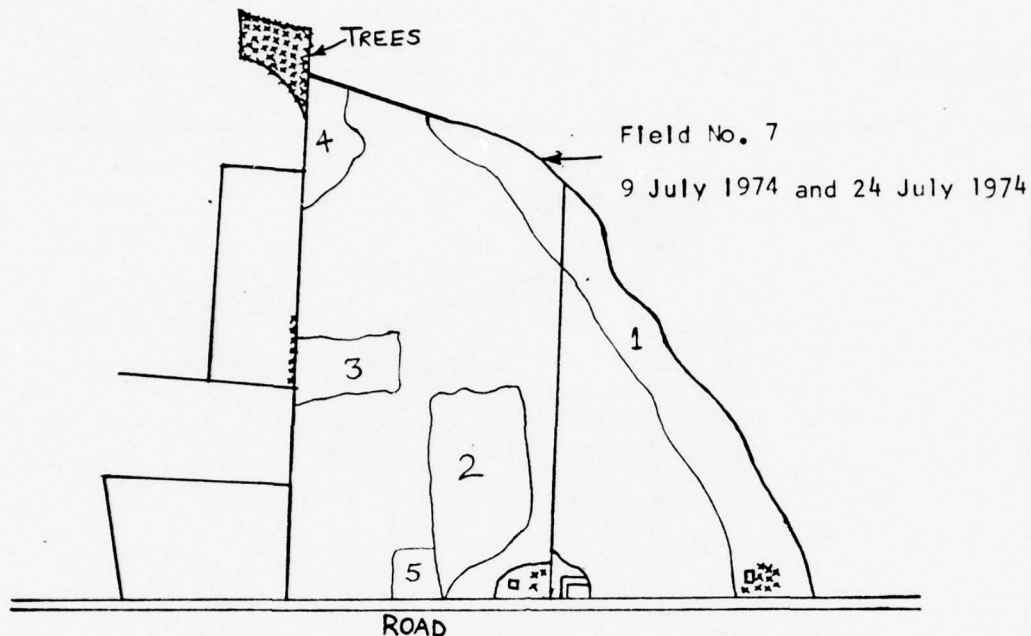
Location	Description
1.	The clay content is the factor influencing the response of the corn. Where the clay content is higher, the corn does not grow as well. Corn height is 40-60 inches.
2.	This area is different from areas adjacent to here due to the fact that there are different hybrids planted. Corn height is 65-70 inches.
3.	This is an area of forming sink holes and probably a slight accumulation of local alluvium.

103

Location	Description
4.	This area is a different soil type; namely, one with a higher coarse fragment content. Also, it is on a slope which may, in part, explain the differences evident on the photography.

24 July 1974

The corn is doing well with no disease or insect damage.

Area 1 (on 9 July sketch map) has a slightly higher clay content which is influencing the corn. The corn is slightly shorter and seems to have been planted later than surrounding areas.

Area 3 is a wet area where the soil is a poorly drained Holly silt loam. During periods of high rainfall, this area is flooded.

Areas 2 and 5 represent two different hybrids planted.

9 August 1974

The corn is doing very well, with the corn in excess of eight feet. There is very minor incidence of Northern Corn Leaf Blight or Stewart's disease, but no other disease or insect damage is apparent. The corn appears to be quite uniform throughout the field.

23 August 1974

Field No. 7

23 August 1974
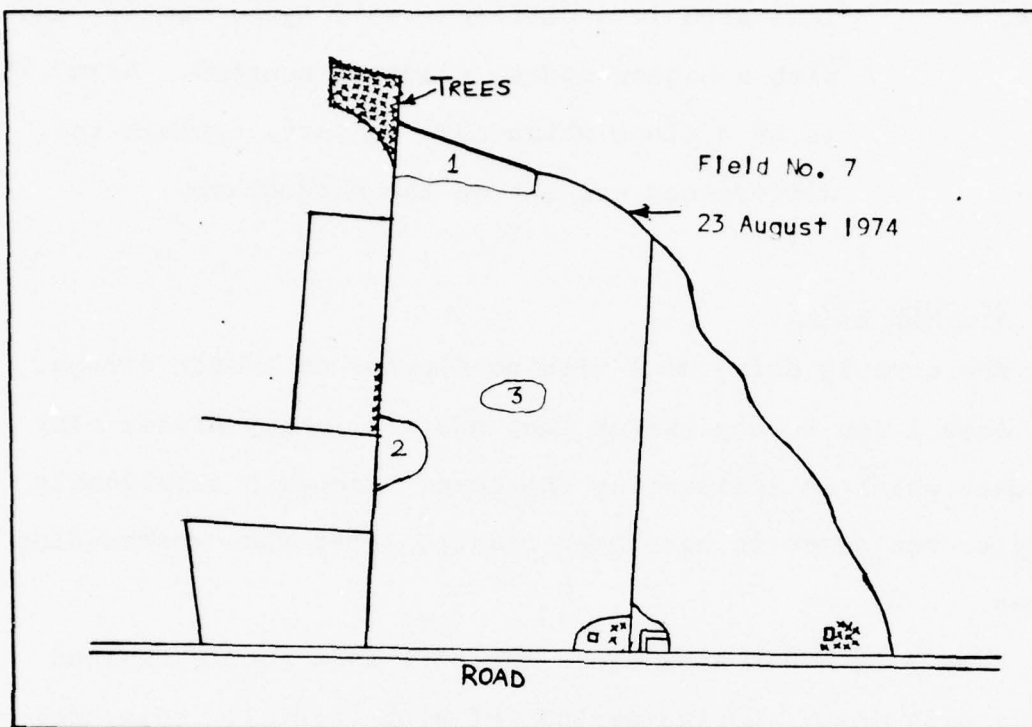
No significant changes; corn doing well.

Area 1 may appear different due to slope or sun angle.

Area 2 is the poorly drained area as noted previously on 24 July.

Area 3 is a depressional area in the field which may be forming a sinkhole.

105

<u>10 September 1974</u>

Predicted yield between 110 and 120 bushels per acre.
Observations by area; intensity of observation is 1 (least)
to 5 (most).

106

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
1.	Average	None	$<$1-1	As planted	7.5-8.5+	110-120

Some yellow leaves ($<$1-1); weeds $<$1-1+; overall condition good.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
2.	High coarse fragment content	None	$<$1-1	As planted	7.5-8.5	85-100

More clay in soil; many more brown leaves than area 1 (.5-1.5); pale green in color; not as healthy looking as area 1; weeds 1.5-2.5.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
3.	Topographic depression	None	$<$ 1-1	As planted	8-9.5	90-100

Depressional area (forming sinkhole); center corn good, (i.e. better than surrounding areas); few brown leaves; weeds 1-2.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
4.	Topographic depression	None	0-1	As planted	7.5-8.5	70-80

Depressional area; corn pale green; weeds $<$ 1-1.

24 September 1974

Estimated yield is between 120 and 130 bushels per acre.

Observation by area; intensity of phenomena observed is on a scale of 1 (least) to 5 (most).

Field No. 7

24 September 1974
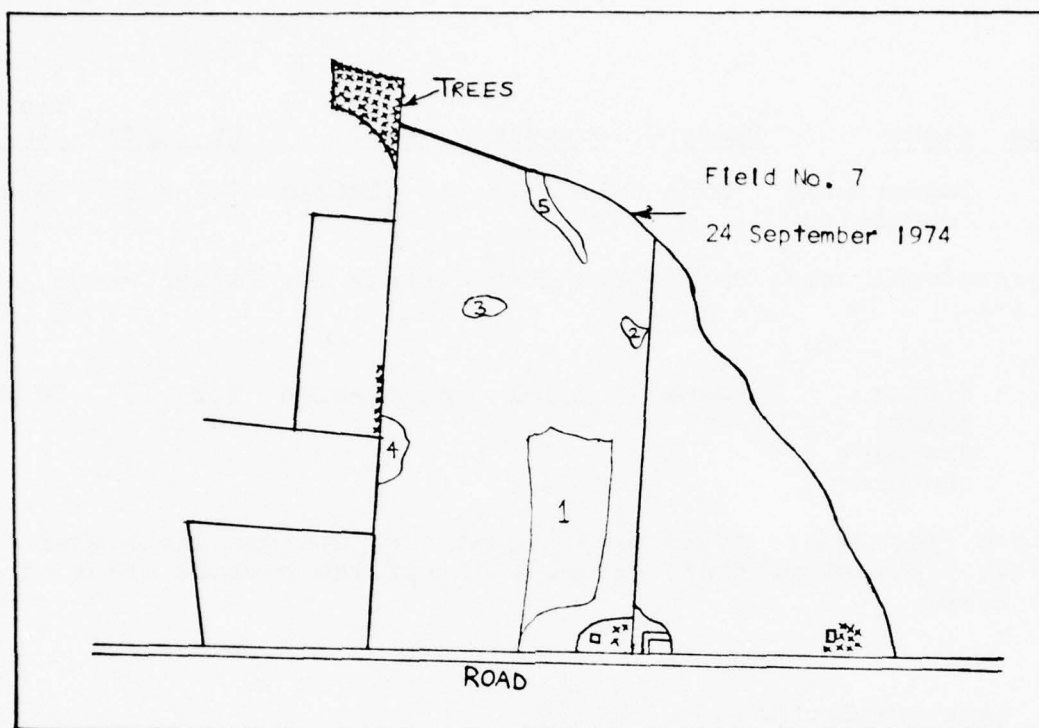
Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
1	Average	None	$<$ 1-1	As planted	7.5-8.5+	120-130

Brown leaves 1-2; weeds $<$ 1-1; corn good.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
2	Higher coarse fragment content	None	$<$ 1-1	As planted	7.5-8.5	100-110

More clay in soil; brown leaves 2-3; weeds 1.5-2.5.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
3	Topographic depression	None	$<$ 1-1	As planted	7.5-8.5	85-95+

Forming sinkhole; brown leaves 1-2; weeds 1-2.

108

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
4	Topographic depression	None	∠1-1	As planted	7.5-8.5	75-85+

Depressional area; corn worse than average for field; weeds ∠ 1-1.

Area	Cause	Insects	Disease	Pop.	Ht. (FT)	Yield BU/A
5	Higher coarse fragment content	None	∠1-1	As planted	7.5-8.5	70-80

Slight ridge where soils have higher clay content and higher coarse fragment content; probably one of the poorest areas of the field.

Late October 1974

Field Seven harvested; average yield 118.2 bushels per acre; range 87.9 to 142.3 bushels per acre.

109

Field Seven Yield Data
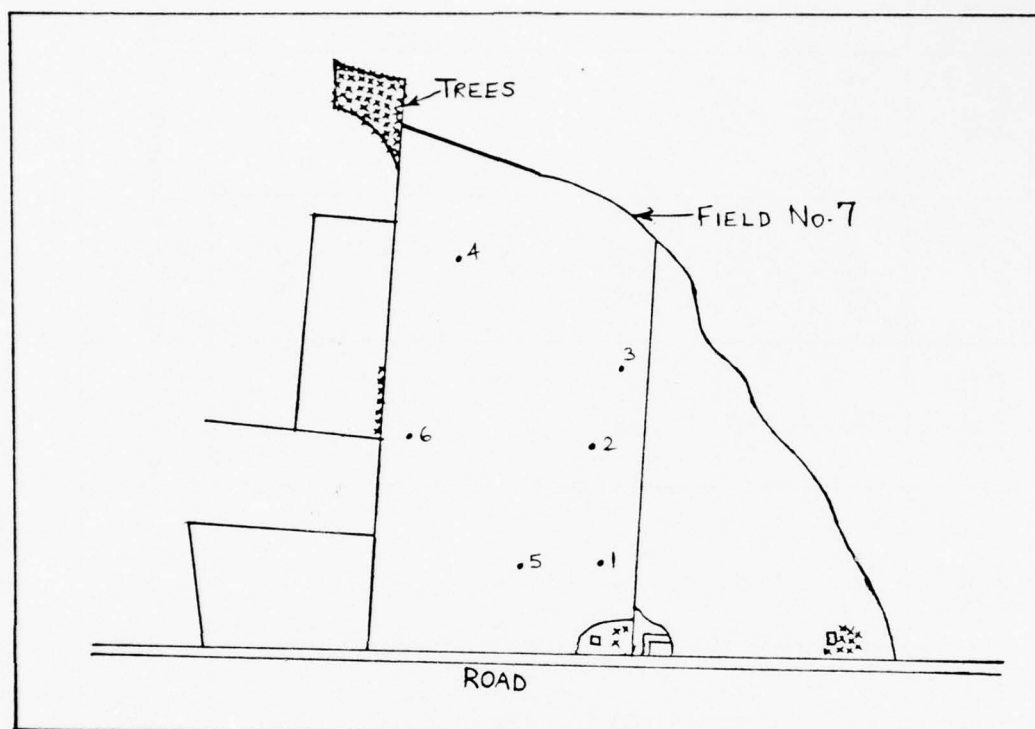
Figure 42. Locations in Field Seven where yield measurements were collected.

Harvest Date:	Late October 1974 (exact date not recorded)
Average Yield:	118.2 bushels per acre
Crops:	Corn, Pioneer 3518 and 3369A varieties
Length of Row Sampled:	27 feet 6 inches
Row Spacing:	38 inches
Sample Area:	1/500 acre

110

Loc.	Gross Wt.	Tare	Net Wt.	# Plants	Down Plants	Barren Plants	Wet Wt.	Dry Wt.	Yield Bu/A
1	39 3/4	14	25 3/4	37	0	2	1128.9	644.2	124.2
2	37 3/4	14	23 3/4	38	0	3	994.2	573.4	115.8
3	36 1/2	14	22 1/2	35	0	2	988.0	563.9	108.6
4	44 3/4	14	30 3/4	40	0	1	1262.0	691.0	142.3
5	41 3/4	14	27 3/4	41	0	1	921.9	511.6	130.2
6	32 3/4	14	18 3/4	34	1	2	934.6	518.4	87.9

Note: Pioneer 3518 variety corn found in locations 1, 2 and 3; Pioneer 3369A variety corn found in locations 4, 5, and 6.

Table 7. Yield measurements at six locations in Field Seven

Significant Results of the Additive Color Analysis of the Multi-
Spectral Photographic Data:

Prior to processing the multispectral data set at the RADC

Image Processing Facility, the positive black and white multi-

spectral transparencies were qualitatively analyzed with an

additive color display (Multispectral Photography Section, pages

54 - 63). Several significant results were determined. The

additive-color study's results and recommendations are presented

here before proceeding with the technicalities of the digital

image processing and the feasibility study in yield prediction

(Reference: Grodewald,20).

General Conclusions:

    1. The set of four narrow band multispectral camera

filters, having the following bandpasses:

> 406 - 480 Nm - Blue
>
> 541 - 581 Nm - Green
>
> 620 - 660 Nm - Red
>
> 810 - 885 Nm - Near-infrared

were selected for agricultural crop stress detection and men-

suration of an agricultural crop. The filters were judged to

have functioned satisfactorily.

    2. Although not specifically selected as corn speciation

filters, it was possible to differentiate corn during mid-July

from a background of normal agricultural crops using the latter

112

spectral bands and additive color display technology (Crop Identification Section, pages 354-361).

3. Weed-infested areas of a corn field displayed significantly different spectral reflectance when compared to noninfested areas of the field.

4. Hydrostressed corn had greater reflectance than normal corn in the 406 - 480 Nm and 620 - 660 Nm bands, the same reflectance in the 541 - 581 Nm band and lower reflectance in the 810 - 885 Nm band.

5. Northern Corn Leaf Blight was detectable because it reduced the reflectance of the plant in the 541 - 581 Nm and 810 - 885 Nm bands while increasing the reflectance in the 620 - 660 Nm band.

6. The spectral signature of corn changes as the growing season progresses. This is due to a combination of factors which include: changing leaf to soil ratio, changing plant configuration (leaf shape, tassel, etc.), and normal changes in the physiology of plants. For this reason, a set of filters selected for corn speciation at a particular portion of the season may not function adequately for these purposes during another portion of the same season.

Digital Image Processing And Multivariate Analysis

for

Crop Yield Prediction

115

Introduction to the Digital Image Processing Phase of the Feasibility Study in Agricultural Crop Yield Prediction

The remainder of this technical report is devoted to explaining the technicalities of this investigation's study of semiautomatic crop stress detection, semiautomatic crop identification, crop yield measurement, and crop yield prediction. Before performing the actual studies, the April to October multispectral photography of Sites One, Three, and Seven required conversion to a digital, temporally-registered, multispectral image format. The image processing techniques used for both converting the multispectral photography to digital imagery and registering the digitized imagery both spectrally and temporally are described in moderate detail. Before proceeding with the discussion of these techniques, the hardware and software capability of the DICIFER Image Processing System, located at the RADC Image Processing Facility, is described in the following section. The technical reports by Forsen(18), Lietz(32), and Zanon(57) provide a more elaborate description of the DICIFER System's hardware and software capability.

RADC Image Processing System Background
Philosophy, and Functions*

RADC has developed a digital image processing capability
within the Reconnaissance and Mapping Branch (IRR) of the
Intelligence and Reconnaissance Division (IR). Full capa-
bility now exists for data handling, preprocessing, searching,
measurement extraction and evaluation, feature data structure
analysis, recognition logic creation, and evaluation of
digitized image data from a variety of sources. Two major
application areas are being actively pursued. The first is
the development of semiautomatic reconnaissance imagery
target-screening and recognition procedures. The second is
land surface thematic mapping based on multispectral images.

Multispectral image data is a set of images collected
over a common ground area where each image represents the
ground information viewed in a different part of the electro-
magnetic spectrum. The recorded spectral information con-
tained in this image set is then processed under the hypothesis
that certain classes of earth surface material can be dis-
tinguished by their spectral signatures.

_____

*Note that this entire section describing the RADC DICIFER
Image Processing System was virtually copied from the RADC
Technical Report, "Spectral Analysis" (Zanon,57) with the
verbal permission of the RADC Project Engineer.

117

RADC Image Processing System Development

The RADC Image Processing System has been commonly known as
IFES/SCORE (Image Feature Extraction System/Spectral Com-
binations for Reconnaissance Exploitation System), based on
titles of the initial software development contracts. Currently,
the system is known as DICIFER (Digital Interactive Complex for
Image Feature Extraction and Recognition), reference Figure 6
on page 15   . The system is interactive to allow the
researcher/analyst the greatest amount of flexibility in guiding
the design of processing logic according to particular
characteristics of the data and output requirements. Imple-
mentation on a general purpose computer has allowed for future
expansion and modification as desired.

The hardware consists of a dedicated mini-computer, two
disks, 7 and 9 track magnetic tape capability, storage display
with hardcopy and line printer. Equipment also exists for
inputting and quantitizing images recorded on such multi-media
materials as opaque prints, film, and multichannel analog tape.
Files can be displayed in black and white (0-255 grey levels)
or pseudocolor, and can be output in the form of color-coded
(64 levels) or black and white transparency files measuring up
to 1024 x 1024 pixels.

The DICIFER software development contracts referred to
previously were directed at establishing the basic operating
system and the applications-oriented measurement routines.

An additional effort provided for the implementation of the
OLPARS (On-Line Pattern Analysis and Recognition System)
capability on the System. OLPARS has been an on-going develop-
ment at RADC and is resident in many forms, both on general-
purpose computing systems and in dedicated configurations
oriented toward the investigation of a particular problem.

The System description given in the following subsections
describes the various processes involved in solving the
types of problems mentioned above. These include three broad
types of applications' routines which perform:

1. Image to image mappings

2. Feature extraction

3. Classification logic design

Table 8 lists specific algorithms on the System. A description
of the hardware and software organization is also included.

To summarize, the DICIFER System provides interactive
general-purpose image processing, feature extraction, and logic
design capabilities necessary to solve a large variety of
problems. Applications of the System to the enhancement of
radar imagery, recognition of tactical targets in aerial
photography, and classification of multispectral data into
land-use categories are being investigated. The results
achieved thus far indicate that rapid progress can be made
toward the solution of a wide class of problems through the use
of this interactive system which contains the necessary
variety of tools.

Image to Image Mappings	Feature Extraction	Classification Logic Design
Smoothing, weighted smooth (includes Laplacian), odd dot elimination, odd line elimination, point edge enhancement, area edge enhancement, max area edge enhancement, thresholding, range change, closed curve routine, Hadamard and Fourier Transforms and inverse trans., filters, grey-level linearization, add and subtract two images, translation, roll picture decimation, interpolation, scale change, fill-in and adaptive fill-in routine, multispectral image manipulation	Spectral coefficients, average grey level, variance of grey level, number of points with specific range of grey levels (includes area and perimeter measurements), connected component measurements, sub-area search executive routines, spectral map extraction, linear combinations, user add-on capability	Pairwise-Fisher logic, piecewise linear logic drawn on (a) optimal discriminant plane, (b) eigenvector, (c) arbitrary vector, (d) coordinate-axis, (e) Fisher direction vector or vector-pair (plane projection); Boolean Compiler to allow logical combination of algebraic expressions to be used in inequalities as decision criteria; Preclustering and Non-linear Mapping algorithm and 1 and 2-space projection for analysis of modality of data classes

Table 8. Capabilities of RADC Image Processing System

The chart, Table 9, provides an overview of the routines available on the System as of this writing, January 1976. This chart shows the frame numbers and titles written as, for example, FRM4 - PREPROCESSING, and the option (routine) number and title written as, for example, under FRM4, 1 - AREA EDIT.

The General Pattern Recognition Problem

Pattern recognition theory grew from an interest in modeling neural behavior and in attempting to imitate by mechanical means the recognition and decision-making functions of man. This interest has not diminished, even though a greater appreciation of what can be accomplished with today's techniques now exists.

The general pattern recognition logic design problem is to create the transfer function of a system that produces the response appropriate to the stimulus. The usual form of response is the generation of a code word identifying the class to which the stimulus is judged to belong. The transfer function is usually a many-to-few mapping with many more stimulus examples than possible classes. The pattern recognition system achieves a partitioning of the sets of example points in a high-dimensional pattern space by means of decision boundaries obtained by design. The system may also be asked to decide when to respond: i.e., when it recognizes the presence of a stimulus to which it should respond. Visual stimuli might be

121

RADC
IMAGE PROCESSING SYSTEM

TABLE 9

FRM 0-MASTER FRAME

1-INPUT
2-DATA EDITING
3-STATISTICS
4-PREPROCESSING
5-SEARCH
6-MEASUREMENT EXT
7-MEASUREMENT EVAL
8-MEAS TRANSFORMATIONS
9-STRUCTURE ANALYSIS
10-CREATE LOGIC
11-FILE MANIPULATIONS
12-UTILITY FUNCTIONS
13-DISPLAY OPTIONS

PREPROCESSING

FRM 1-INPUT

1 - SDS INPUT-U
2 - SDS INPUT-D
3 - COMPUTER SPEC INPUT
4 - DISK TO TAPE
5 - TAPE TO DISK
6 - TAPE DIRECTORY
7 - CHANGED TAPE
8 - INITIALIZE TAPE
9 - LIPS TAPE INPUT
10 - COLOR PRINTER
11 - SINGLE IMAGE
12 - DISPLAY OPTIONS
13 - MASTER FRAME

FRM 3-STATISTICS

1-DENSITY
2-MIN DIF
3-MAX DIF
4-AVE DIF
5-DENSITY-AREA
6-MIN DIFF-AREA
7-MAX DIFF-AREA
8-AVE DIFF-AREA
9-PERCENT
10-SPECTRAL MAP
11-COMP SPEC MAP
12-MULTI COMP SPEC MAP
13-BINARY COMP SPEC MAP
14-MISC STATISTICS
15-SINGLE IMAGE
16-DISPLAY OPTIONS
17-MASTER FRAME

FRM 2-DATA EDITING

1-RECT AREA SELECT
2-RECT AREA SEL-SPEC
3-CURSCR RECT AREA SEL
4-CUT RECT AREA SEL-SPEC
5-DRAW AREA BOUNDS
6-BLOW UP DRAW AREA
7-AREA COORDS
8-BOX AREA
9-PICK REG POINT
10-REG IMAGES
11-IMAGE REGIST
12-DISPLAY X-Y
13-BLOW UP IMAGE
14-IMAGE TO SPEC
15-SPEC TO IMAGE
16-INIT DSGN VECT FILE
17-INIT TEST VECT FILE
18-SINGLE IMAGE
19-DISPLAY OPTIONS
20-MASTER FRAME

FRM 4-PREPROCESSING

1-AREA EDIT
2-THRESHOLD
3-NORMALIZE
4-ELEMENT CHANGE
5-RANGE CHANGE
6-SHIFT PIC
7-DECIMATION
8-SMOOTH
9-NOISE ELIMINATION
10-LINE MANIPS
11-COMBINE IMAGES
12-TRANSFORMS
13-SINGLE IMAGE
14-DISPLAY OPTIONS
15-MASTER FRAME

FRM 15-SMOOTH

1-BOX SMOOTH
2-NBM SMOOTH
3-WEIGHTED SMOOTH
4-WGHTD SMOOTH-ABSOLUTE VAL
5-SINGLE IMAGE
6-DISPLAY OPTIONS
7-RETURN

FRM 17-COMBINE IMAGES

1-ADD IMAGES
2-DIFFERENCE
3-ABSOLUTE DIFFERENCE
4-DIFF FACTOR
5-RATIO IMAGES
6-SPECTRAL SIGNATURE
7-SINGLE IMAGE
8-DISPLAY OPTIONS
9-RETURN

FRM 16-NOISE ELIMINATION

1-ODD DOT ELIMINATION
2-ODD LINE ELIMINATION
3-SINGLE IMAGE
4-DISPLAY OPTIONS
5-RETURN

FRM 19-TRANSFORMS

1-HADAMARD TRANSF
2-HADAMARD FILTER
3-FOURIER TRANSF.
4-INV. FOURIER TRANSF
5-SINGLE IMAGE
6-DISPLAY OPTIONS
7-RETURN

FRM 18-LINE MANIPULATIONS

1 - POINT EDGE DETECTION
2 - AREA EDGE DETECTION
3 - AREA EDGE DETECTION-MAX
4 - SIMPLE FILL IN
5 - ADAPTIVE FILL IN
6 - CHANGE LINE
7 - EDGE PACK
8 - SYM EDGE DISPLAY
9 - GRID - DENSITY MAP ADDS
10 - SINGLE IMAGE
11 - DISPLAY OPTIONS
12 - RETURN

FRM 20-MISC STATISTICS

1-LOCAL STATISTICS
2-DETAILED LOCAL STATS
3-DISTANCE
4-DIST-D
5-DIST-U
6-POINT CNT
7-IMAGE MASK
8-SINGLE IMAGE
9-DISPLAY OPTIONS
10-RETURN

MEASUREMENT EXTRACTION

FRM 5-SEARCH

1-RECTANGULAR SEARCH
2-SINGLE IMAGE
3-DISPLAY OPTIONS
4-MASTER FRAME

FRM 7-MEASUREMENT EVAL

1-SELECT DATA SET
2-PAIRWISE HISTOGRAMS
3-MULTI HISTOGRAMS
4-PROB OF CONF
5-DISCRIMINANT
6-SINGLE IMAGE
7-DISPLAY OPTIONS
8-MASTER FRAME

FRM 6-MEASUREMENT EXT

1-SPECTRAL MEAS-DSGN SET
2-SPECTRAL MEAS-TEST SET
3-CREATE MEAS. SET
4-MEAS. TO VECTOR
5-ADD MEAS. TO VECTORS
6-OUTPUT VECTORS
7-SINGLE IMAGE
8-DISPLAY OPTIONS
9-MASTER FRAME

FRM 8-MEAS TRANSFORMATIONS

1-SELECT DATA SET
2-MEAS TRANSFORM-DESIGN SET
3-MEAS TRANSFORM-TEST SET
4-SINGLE IMAGE
5-DISPLAY OPTIONS
6-MASTER FRAME

STRUCTURE ANALYSIS & LOGIC DESIGN

FRM 9-STRUCTURE ANALYSIS

1-SELECT DATA SET
2-ONE SPACE-HISTOGRAM
3-TWO SPACE-CLUSTER PLOT
4-POPULATION MAP
5-SINGLE IMAGE
6-DISPLAY OPTIONS
7-MASTER FRAME

FRM 10-CREATE LOGIC

1-SELECT DATA SET
2-SELECT NODE
3-DISPLAY CUR TREE
4-CORRELATION MATRIX
5-CREATE FISHER LOG
6-CREATE 1-SPACE LOG
7-CREATE 2-SPACE LOG
8-CREATE BOOLEAN LOG
9-LOGIC EVAL
10-EVAL TEST SET
11-DELETE LOGIC
12-GET MEAN AND COV
13-EDIT LOGIC
14-CORRELATION MATRIX
15-DECISION DISP
16-DECISION IMAGE
17-LOGIC OUTPUT
18-SINGLE IMAGE
19-DISPLAY OPTIONS
20-MASTER FRAME

AUXILIARY ROUTINES

FRM 11-FILE MANIPULATIONS

1-DELETE FILES
2-RENAME FILES
3-DIRECTORY LISTING
4-USER PROG. DIR.
5-INITIALIZE
6-ADD USER FUNCTION
7-DELETE USER FUNCTION
8-SINGLE IMAGE
9-DISPLAY OPTIONS
10-MASTER FRAME

FRM 13-DISPLAY OPTIONS

1-SINGLE IMAGE-DIST
2- -UNDIST
3- -INTERP
4- -BLOW UP
5-SIDE BY SIDE-DIST
6- -UNDIS
7- -INTERP
8- -BLOW UP
9-TOP AND BOTTOM-DIST
10- -UNDIST
11- -INTERP
12- -BLOW UP
13-DISPLAY AREA
14-AREA LINES
15-AREA LINES-IMAGE
16-CLEAR GRAPHICS
17-FLICKER DISPLAY
18-BINARY FLICKER
19-DISPLAY UTILITY FUNCTS
20-RETURN

FRM 12-UTILITY FUNCTIONS

1 - DIR
2 - ADD LOG
3 - SINGLE LINE HISTO
4 - MAKE STC
5 - CORRELATE TEST FILE
6 - TEST SDS-RDS DISK
7 - DIRECTORY AID
8 - OCTAL BYTE
9 - DECIMAL RDS-UNSIGNED
10 - DECIMAL RDS-SIGNED
11 - DECIMAL BYTE
12 - UNARAY DUMP
13 - ANALYSIS INDEX
14 - SCALE IMAGE
15 - FRACTAL DIMENSION
16 - MASTER FRAME

FRM 14-DISPLAY UTILITY FUNCTIONS

1-PICTURE ON
2-PICTURE OFF
3-GRAPHICS ON
4-GRAPHICS OFF
5-NEG PIC
6-ROLL
7-PICTURE PRINT
8-GRID-U
9-GRID-D
10-RETURN

122

present on a page of printed text, an aerial photograph, a set

of spectral images, or a photomicrograph of biological cells.

Images, or portions thereof, are usually transduced into an

equivalent electronic form by digitizing discrete samples

spaced at regular intervals for subsequent storage and proces-

sing. A search function has the task of isolating the per-

tinent information which is to be recognized.

The process of generating responses usually involves a

sequence of concatenated operations, including preprocessing

and feature extraction, and classification, as well as the raw

data transduction and search operations already mentioned.

Image-to-image preprocessing transformations are sometimes

applied to correct for known systematic geometric and radio-

metric distortions, to filter out redundant data, to enhance

certain information for visual presentation and/or to trans-

form the data to make it easier to extract features. These

features (or measurement values) are used by the classification

logic to accomplish the recognition task if object (or point)

classification is desired.

The derivation of a useful set of features is the most

critical part of the solution to any pattern recognition

problem. The ideal selection criterion is that they possess

only the essential information which adequately differentiates

objects according to class membership. A set of features for

an object (or point) is normally referred to as a feature

123

vector, and it is this vector which is passed on to the classification logic to be recognized as belonging to a class set. The classification logic determines the sub-space in which the feature vector is located, and generates the class decision code along with an indication of the confidence in that decision.

The choice of features and of classification logic for a given problem is obtained at present largely by an ad hoc procedure. Feature definition, hence the structure of the sets of feature vectors, is strongly dependent on the application, as well as on prior transformations. Satisfactory solutions to difficult problems seem to require many hours of study and experimentation with large amounts of representative data and a thorough understanding of the underlying physical phenomena which dictate the empirically observable class variations. Furthermore, the tools of the analyst often determine the ease with which a solution is obtained and probably the nature of the solution as well. Figure 43 is a flowchart of the typical process for solving pattern recognition problems.

DICIFER's Image Processing System Philosophy

The RADC Image Processing System is intended to be used as a research tool in a highly interactive manner. A researcher would typically try many different algorithm sequences with

124
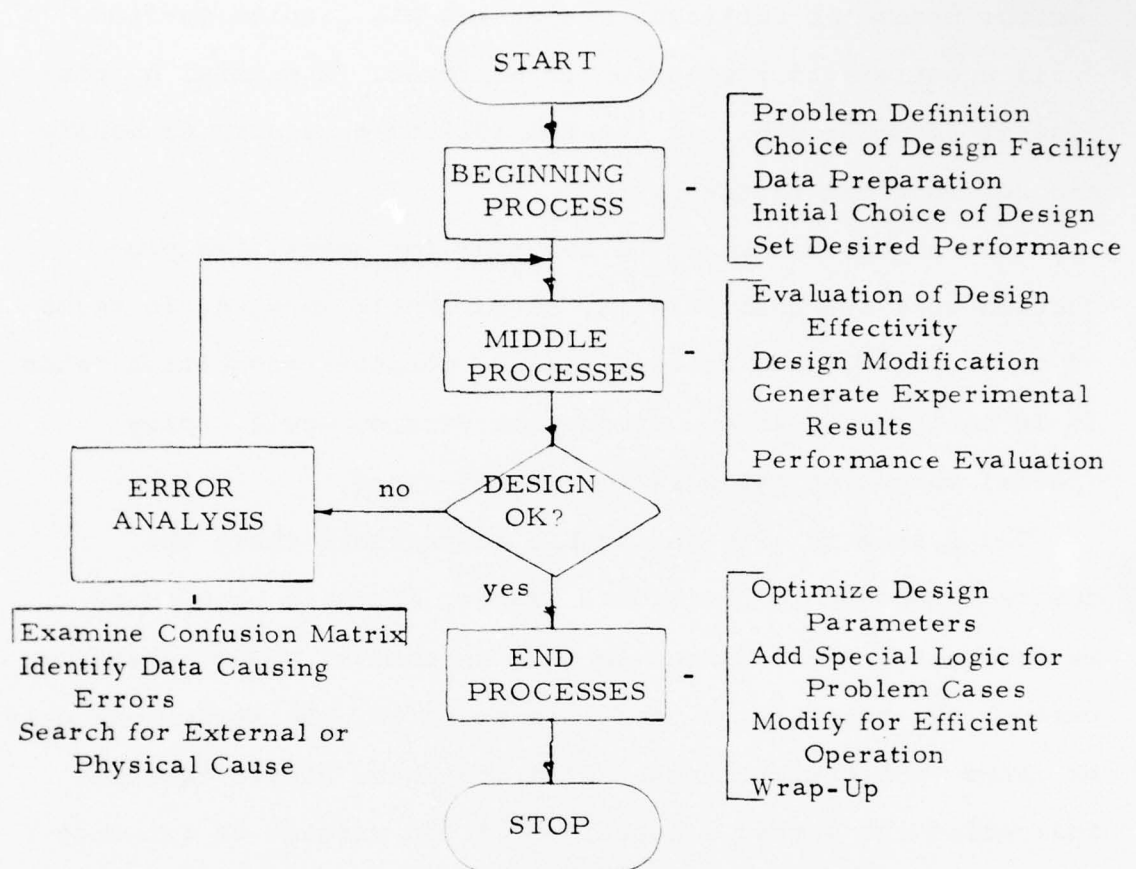
Figure **43**

Typical Process for Solving Pattern
Recognition Problems

various parameter settings, evaluating his results on-line until a satisfactory solution is achieved. A general purpose digital computer provides the flexibility necessary to modify and add routines to the System.

Some algorithms would be too slow for repetitive production work but their utility is currently measured in terms of other performance criteria; e.g., correct recognition rates. It is anticipated that a production version would employ special computing hardware to enhance speed.

The System is also applicable to problems where the desired response is additional imagery in which objects to be recognized are enhanced so that an analyst may more easily perform the recognition task. In addition, the System may only be asked to highlight those areas that have simple features indicating the probable locations of the targets of interest and then provide the figure-background separation necessary for object identification.

The methodology used to obtain solutions to pattern classification problems requires that these capabilities reside in one system if convergence to a solution is to be efficiently achieved. It is only through the application of classification logic that the adequacy of a feature set can be ascertained. It is through the process of error analysis that ideas for modifications and additions to the feature set, to improve discrimination, are generated. When the modifications and additions have been made, new logic must be

126

designed and tested, new errors investigated, etc. The System
was designed to accommodate this cycle of feature design, logic
design, and error analysis.

Facility Hardware

The hardware system is configured about the following
primary components: DEC (Digital Equipment Corporation) PDP-
11/20 Computer System, SDS Computer Eye Model 108-3A Input and
Display System, Tektronix Interactive Terminal, and RADC Color
Output Film Printer. Brief descriptions of each component and
its function are contained in the following paragraphs.

PDP-11/20 Computer System

The PDP-11/20 Minicomputer performs all processing, file
manipulation, and input/output functions. The system consists
of a PDP-11/20 processor with 28,672 16-bit words of core
memory (4096 locations are assigned to the Unibus). An
Extended Arithmetic Element, RS-11 256K word fixed head disk,
RP02 10 million word disk pack unit, DEC seven and nine track
industry-compatible magnetic tape unit, dual DEC tape unit,
teletype and card reader. The 10 million word disk pack is
used mainly for storage of digitized images. The system
routines are stored on the fixed head disk and are called into
core by a resident executive program in response to user
requests entered via the keyboard. User communication with

the system is via the Tektronix display terminal, teletype, and/or cursor and monitor.

SDS Computer Eye Model 108-3A Input and Display System

The SDS Computer Eye Model 108-3A provides for both the input of photographic imagery data through a special vidicon-type computer television camera and the resultant image display in black & white and/or color on two CRT monitors. An image may be displayed by the Computer Eye camera prior to digitization or the image's digital equivalent may be displayed. In the display mode, the system accepts a digital image file from the PDP-11/20 and displays it in color and/or black and white using internal disk refresh capability. The SDS Computer Eye Model 108-3A is composed of the vidicon scanner, displays, the 108 digitizer chassis, and a self-contained controller to interface the Computer Eye to the PDP-11/20 Minicomputer. A 804-2 joystick-controlled cursor is utilized for interactive image processing. In addition, the current digitizing system includes a digital video convertor which provides digital-analog interface for the color television displays, and a Data Color 703 Color Display Controller, which provides the color display with 32 color-coded levels for density slicing. The black and white Miratel monitor and refresh storage device, which together displays the Computer Eye's digitized image, 512 x 512 picture elements (pixels),with 256 grey level codes,

128

are interfaced to the DICIFER system and provide the primary
hardware for interactive image processing. Raw video, digitized
images, cursor cross or any combination can be displayed
separately or superimposed on the displays.

The Tektronix Interactive Terminal

The Tektronix Interactive Terminal is a prime instrument of
the man-machine interface; it displays software menus and
system dialogue as well as graphics which aid in analysis
(e.g., histograms and scatterplots). Any information displayed
can be retained in hardcopy form. The hardware consists of a
Tektronix 4010-1 storage tube display terminal and a 4610 heat
processing hardcopy output unit.

The RADC Color Printer

The RADC Color Printer is a modified (in-house) film
printing device which allows output printing of digital image
files from the PDP-11/20 system. This device utilizes a rotat-
ing scan mechanism to expose color film on a pixel-by-pixel
basis. Output image files can be as large as 1024 by 1024
pixels. The 256 original grey levels are mapped into 64
distinct colors. The output image (1024 x 1024) is recorded
on a 4.5 by 4.5 inch area on 5-inch wide color film. The
output images exhibit excellent color.repeatability and geo-
metric fidelity.

Figure 44 depicts the physical configuration of the complete DICIFER System.

Software

There are three modules of applications software:

(1)   Preprocessing

(2)   Measurement Extraction

(3)   Structure Analysis & Logic Design

These programs are called by the executive routine according to user selection from option menus. They depend heavily on various display routines which allow the analyst to view imagery and view the results of algorithms which operate on imagery and/or vector data. Figure 45 depicts the basic block diagram of the system. Convenient file manipulation and logical executive control are important aspects of the System operation. The following briefly describes the design algorithms available.

Preprocessing

In the preprocessing phase, the image is processed to enhance the characteristics of the objects to be recognized or, conversely, to eliminate information from the image not pertinent to the detection and classification of the objects. The output is another image. Smoothing, noise elimination, edge enhancement, and "line manipulation" algorithms (which operate on edge-detected images) comprise the "local" mapping

Figure 44. RADC Image Processing System

131

EXECUTIVE

FILING SYSTEM

PREPROCESSING

MEASUREMENT EXTRACTION

STRUCTURE ANALYSIS & LOGIC DESIGN

AUXILIARY ROUTINES

Smoothing	Create Area	Structure	Display
Noise Removal	Bounds	Analysis	Input/Output
Edge Detection	Create Vector	Logic Creation	File Manip.
Line Manip.	File	and Eval.	Data Editing
Transforms	Measurement		Statistics
Grey Level Mod.	Eval.		Utility Routines
Image Restoration	Measurement		
Image Combination	Transf.		

Figure 45.   BUILDING BLOCKS OF SOFTWARE SYSTEM

in the preprocessing module. These are local mappings in the sense that the grey level at a point in the preprocessed image is a function only of grey levels at points in the neighborhood of the corresponding point in the original image. This distinguishes them from the Fourier and Hadamard transformations and filters which are global in nature and which are also contained in this module. Because the output of any of the preprocessing algorithms is in image format, many preprocessing options may be concatenated by using the output image of one algorithm as the input image to the next algorithm, etc., until the desired result (enhancement) is achieved.

Measurement Extraction

The **search** routines are applicable when the target object is small in relation to the size of the image. These translate a rectangle around the image applying specific search criteria to picture elements in the rectangle. A criterion may be based on masking operations, texture measurements and/or topological measurements. The positions of the rectangle which satisfy the criterion are noted and the enclosed "areas of interest" define those area files (sub-region boundaries) which are callable for further processing.

The search function serves to prescreen the data, eliminating with high reliability those portions of the image which do not contain the target object while locating areas which

133

may or may not contain target objects. For example, if the
problem were to find and classify aircraft in aerial photo-
graphy, a search function, which operates on an edge-detected
image and simply finds rectangles which contain a sufficient
number of connected edge points, would eliminate most of the
image from further processing.

In the feature (measurement) extraction phase, measure-
ments (features) are taken in the areas of interest found by
the searching algorithm or in user-drawn sub-areas. These
features can be based on the same properties of texture, shape
and topological characteristics upon which a search algorithm
may be based. However, they would, in general, be too compli-
cated to merit application to the total image area. The set of
features must provide the distinction between targets and tar-
get facsimiles as well as the distinction between different
classes of targets. Because of their application dependence,
provision is made for user-supplied routines for extracting
features. Examples of features that may prove useful include
Hadamard coefficients: grey-level spatial dependency coeffi-
cients and topological measurements applicable to binary (edge-
detected) images consisting of the number of connected com-
ponents, number of points in the largest connected component,
etc. Methods of encoding and measuring the shape of a portion
of a binary image, such as chain encoding, have been left, as
with the others, to the analyst to define.

134

The output of all measurement or feature extraction routines is a Vector File. This file contains an L-dimensional vector of L or M measurement values, and ancillary data such as desired response (class code), for each item (object or point) to be recognized. These data are used by the Structure Analysis, Logic Design and Measurement Evaluation Routines.

The user is provided two methods for evaluating the discriminatory value of each measurement. In essence, they both provide means for selecting a subset of measurements. The measurements which are chosen for retention define the coordinate subspace and the desired projection to a lower-dimensional space.

An optimal method for selecting a subset of M measurements must consider the decision logic criterion, such as the Bayes Risk or the probability of error. This, in turn, requires the estimation of the joint probability functions for all possible n-tuples. The computational difficulties in obtaining such an optimal ranking preclude this approach in all but the simplest problems. Therefore, the sub-optimal algorithms of the Discriminant Measure and Probability of Confusion Measure are provided as options to rank order the L measurements x_1, x_2, ..., x_L. Each algorithm provides three distinct types of rankings. The first uses a significance measure of a particular component, say x_p, for discriminating class i from class j. This significance will be designated by $M_{ij}(x_p)$. The second type of ranking uses a significance measure of x_p for

135

discriminating class i from all other classes, and is desig-
nated $M_i(x_p)$. The last type of ranking uses a measure of the
overall significance of x_p for discriminating all classes and
is designated $M(x_p)$.

Structure Analysis, Logic Design, and Evaluation

In the structure analysis phase of the logic design, the
analyst attempts to determine how the vectors from the dif-
ferent classes are distributed in the L-dimensional feature
space.

Parametric techniques of pattern recognition theory assume
that the data from each class are unimodal and distributed
according to a particular probability distribution, such as
multivariate Gaussian. In real world problems, however, such
assumptions are risky and may lead to poor results.

Structure analysis assists the analyst to learn the modality
of each class. The vectors which belong to a mode (single class
cluster) can be relabelled and treated as subclasses which are
later recombined by the decision logic.

Projection of the data onto the two-space spanned by those
two eigenvectors of the lumped covariance matrix which have the
largest eigenvalues, is one classical method for determining
structure of multidimensional data. A mode can be separated
from the rest of the vectors by drawing a piecewise-linear,
convex boundary on the display around it, and specifying a new

136

label for the vectors represented by the points within that boundary. Similar routines exist for projecting histograms on a single vector. The one- or two-space vectors may also be arbitrary, original feature coordinates, or Fisher vectors in addition to the eigenvectors. For example, an "Optimal Discriminant Plane" projection projects points onto first a Fisher direction (for a chosen pair of classes) and then onto the Fisher direction orthogonal to the original.

The outlines of the boundaries from which these vectors were extracted may also be displayed against the background of the image from which they came, thus allowing the analyst to see the specific areas of the image which comprise the particular mode. This may allow him to perceive the _physical_ reason for the mode to exist. The desire to be able to relate regions of the L-dimensional feature space back to the original source information was the primary motivation for adding OLPARS capabilities to the DICIFER System.

A powerful non-classical technique for structure analysis included in the System is the Non-Linear Mapping algorithm. The algorithm calculates a mapping of points in the L-dimensional feature space to a two-dimensional space which attempts to preserve interpoint distances. It includes a preclustering routine to reduce the number of points to manageable levels.

The decision logic for implementing the desired classification must be based on the set of vectors available at the time of design. If one is fortunate enough to have continual

137

access to new data, it is desirable to be able to update the decision logic if necessary. The redesign may affect the entire logic of some types of classifiers, which may require substantial effort to accomplish. If performance does not extrapolate well to a particular pair of classes, it is convenient to be able to specialize in upgrading their responses without affecting other classes.

A hierarchical structure is used in the DICIFER System to allow easier upgrading of the classifier when needed. It also allows the total problem to be divided up into smaller problems which are, perhaps, easier to solve. This structure also follows naturally from the fact that many classification schemes developed in the past have specialized in two-class dichotomies. The process of obtaining a response to a given vector input can and often does involve a sequence of partial decisions. Thus, the structure of the decision logic is tree-like with branches and nodes. The Logic Creation and Evaluation module includes, but is not limited to, the following types of logic which the analyst may call upon for use at any node of the decision tree: Fisher Pairwise, One- and Two-space projections, and Boolean. The system keeps a record of the decision logic as it is being designed. The record becomes the classification procedure for the recognition of the new data.

Fisher discriminant logic first calculates the Fisher direction vector $\vec{d}_{ij}$, and a threshold θ_{ij} for each pair of classes A_i and A_j in the data base.

Classification of an unknown vector $\vec{V}$ is done on a pairwise basis. For each pair of classes A_i, A_j, the following decision procedure is followed: If $\vec{V} \cdot \vec{d}_{ij} \geq \theta_{ij}$ a vote counter for class A_i is incremented. If $\vec{V} \cdot \vec{d}_{ij} < \theta_{ij}$ a vote counter for A_j is incremented. When all pairwise tests have been completed, the class with the maximum number of votes is taken as the decision. If two or more classes are tied with the maximal number of votes, the vector $\vec{V}$ is assigned to a reject class R as its decision class. Several other reject criteria are also possible.

There are several kinds of plane projection logics, but all require that two vectors in the L-dimensional space be chosen. They may be two eigenvectors of the lumped covariance matrix, the vectors from the Optimal Discriminant Plane, a pair of Fisher directions, two coordinate axes, or arbitrary vectors designated by the user. The data are projected on each vector and the pair of values determine coordinates of points in a two-dimensional plane displayed on the CRT; the identity of each vector being indicated by the class symbol of that vector. The analyst may now draw (by designating end points of line segments with the aid of a cursor) several piecewise linear boundaries to separate classes or groups of classes from

139

one another on the display. The analyst may also designate a region as a reject region. The logic will classify an unknown vector $\vec{V}$ by projecting it onto the plane and determining into which region it falls. The decision at this point may be only a partial decision because such a region may contain more than one class. Logic can also be designed using projections on single vectors.

The Boolean Logic option allows the analyst to write decision criteria in the form of Boolean algebraic statements on the coordinates of the feature vector. Any meaningful statement involving the algebraic operations of addition, subtraction, multiplication or division; integer constants; the equality and inequality symbols; or the logical connectives conjunction ($\wedge$) and disjunction ($\vee$) may be written. A dichotomy is achieved based on whether the predicate is true or false.

A feature of the System (not found in previous OLPARS installations) is the ability to evaluate the logic at each node before proceeding with other node logic designs. Formerly, the entire decision tree had to be completed before evaluation could take place. It was found important to allow the analyst to detect and correct deficiencies as they were made.

A functional overview of the software system is presented in Figure 46. Within the function boxes there is a number (n) which corresponds to the frame number (s) concerned with that

140
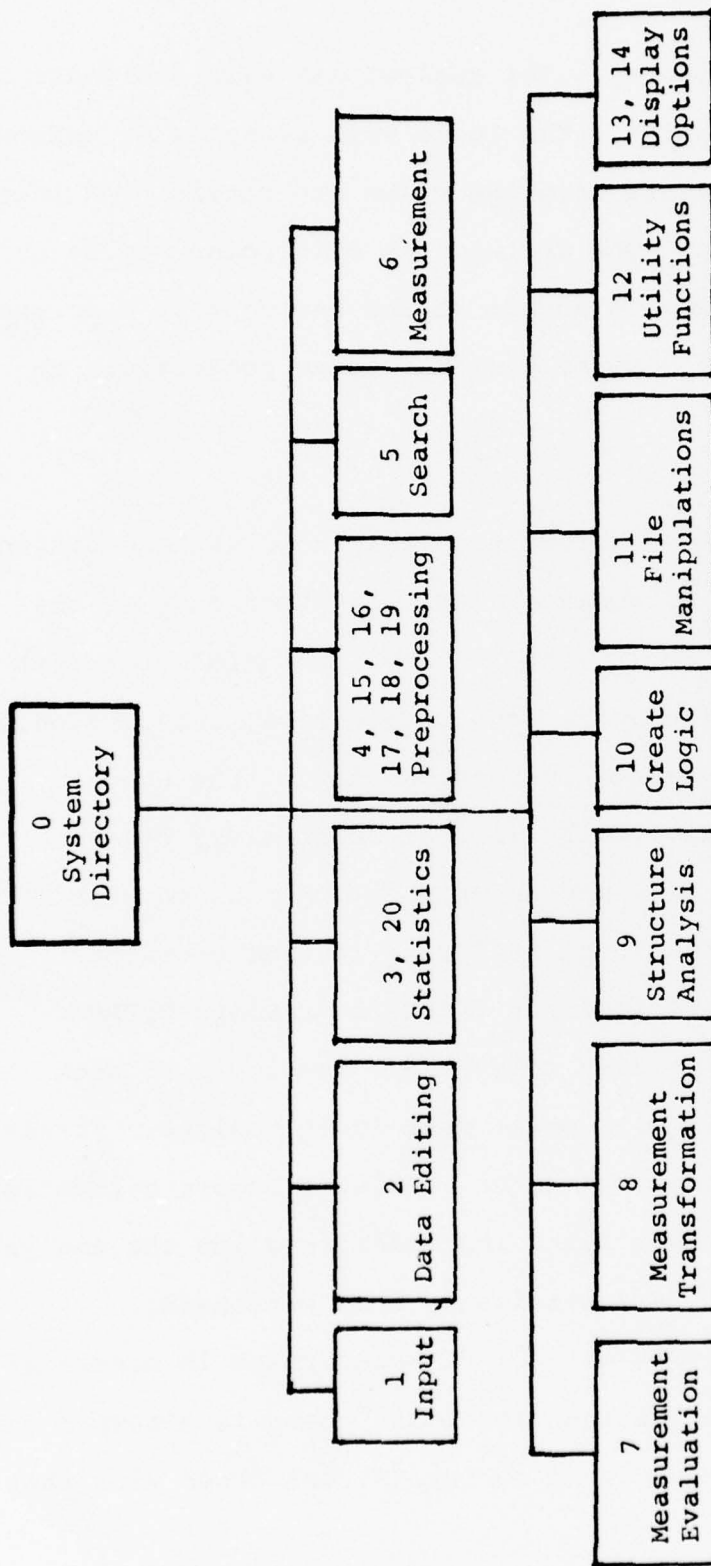
Figure 46. System Functional Overview

141

function. The software system consists of thirteen functional modules, each of which consists of one or more frames.

The following is a brief description of these functions:

Input: This portion of the system contains routines which control the scanning and digitization of an image by the SDS Computer Eye System. Digitized images may also be read to and from magnetic tape, and transferred to a color printer for hardcopy output.

Data Editing: This module consists of routines to specify sub-areas of images for later processing, to register images, and to transfer data from one file structure to another.

Image Statistics: This module allows the user to compute basic statistics on all or part of an image. The routines in this module calculate several types of distributions of the grey-level information contained in an image. Histogram and/or line graph displays of these distributions are useful in the determination of optimal settings of parameters for the line enhancement options as well as in the determination of optimum grey-level values for the Range Change option.

Preprocessing: In this module the user is given a choice of several preprocessing operations. Each preprocessing function alters in some manner the image specified by the user. The most common type of preprogrammed callable preprocessing operations are those in which each point in the raster is examined and may possibly have its grey-scale value changed,

142

depending upon its current value and/or the values of the other points in some neighborhood around it. After one point has been processed, the adjacent point on the right hand side, or if this has been the last point in a line, the first point in the next line, will be processed. The pattern is continued until the last point in the last line has been processed.

Examples of preprocessing routines of this kind are the various smoothing, edge-enhancement and noise elimination routines. Another class of preprocessing routines is the Combine Image routines. These add or subtract two images on a point-by-point basis.

Search: The routines in the search module attempt to emulate the human eye in that they scan images, applying various searching algorithms as specified by the user in order to (a) pinpoint the positions of the points of interest or (b) find the boundaries of regions of interest. The former would normally consist of center points of tactical objects (e.g., ships, trucks) while the latter would be regions in which particular tactical objects may be found (e.g., water bodies, roads).

Measurement Extraction: The routines in this portion of the System extract a set of scalar values from each of the neighborhoods found by the searching operations, or on a point basis in the case of spectral imagery. The set of measurement functions offered to the user includes routines to measure the

average density, range of densities, low density, high density, median density, and the deviation of the densities within a neighborhood. Furthermore, it is possible for the user to call upon one or more measurement routines of his own design by calling a user program.

Measurement Evaluation: This portion of the System provides both displays and computational routines which give an indication of the utility of the measurements. The ultimate verdict depends upon the success of the logic which has been designed using these measurements. "If the results are good, the measurements are good". However, this phase will point out totally inadequate measurement sets before the user has had to go through logic design and classification. It will also point out specific weaknesses. For example, the measurement might be good for discriminating class A from class B but ineffective in discriminating class B from class C.

Measurement Transformation: This module provides the capability to generate new vector sets through algebraic combination of the measurements in an existing set.

Structure Analysis: This phase allows the user to examine how the data vectors are distributed in the measurement space with the particular purpose of detecting multimodal data. One- and Two-Space projections and a Nonlinear Mapping capability are provided as tools.

144

Logic Creation and Evaluation: This phase contains a large set of logic creation algorithms which do not depend upon any specific assumptions regarding the probability distributions of the measurements. The types of logic include Fisher logic, One- and Two-Space logics, and Boolean logic. Several methods are provided to evaluate the logic following its creation.

File Manipulations: This module provides the basic functions including deleting a file, renaming a file, and listing the file directory. This module also allows the user to input and run routines he has programmed.

Utility Functions: This module contains routines not directly applicable to the solution of pattern recognition problems. It is intended primarily as an aid in debugging new routines by providing various disk dump and file checking options.

145

Introduction to Digitization

To investigate the feasibility of agricultural crop yield prediction utilizing multivariate and regression analyses of temporally-registered, narrowband multispectral imagery, required the analogue to digital conversion (digitization) of the multispectral photography to a proper digital format. Since the study was planned to demonstrate this feasibility of crop yield prediction through analysis of the fluctuation in a crop's multispectral albedo (reflectance) over the course of the growth cycle, the digitized set of multispectral images would have to be registered both spectrally and temporally. This meant that the digital image of an arbitrary field of corn would graphically correspond, on a pixel (picture element) for pixel basis, to the other digitized multispectral images of the corn field (reference Figure 64 , page 187).

The initial step in the development of a temporally-registered, multispectral image data set was the digitization of each of the four 55mm by 105mm (2 inch by 4 inch) spectral scenes, which were imaged in each nine inch frame of multi-spectral photography. Following digitization, each digital spectral image was registered to the other digital spectral images originating from the same frame of photography. Each month's digitized and registered multispectral images must then be registered with the images from the other months'

146

digitized and spectrally registered images. Finally, the

digitized and registered multispectral imagery must be

preprocessed to minimize the negative effects resulting

from any missed photo coverage over an area of the selected

agricultural test site (holidays) which may have occurred

due to cloud cover, pilot error, or camera cycling. This

section will describe the techniques employed during the

digitization of the multispectral photography of the three

useable agricultural test fields, Sites One, Three, and Seven.

Digitization Procedures

Prior to the digitization of any multispectral photography

of Sites One, Three, or Seven, several critical reviews of

the total photography data set were performed. After selecting

the useable data of the three sites, each frame of multi-

spectral photography was investigated. The photography which

had too many clouds or holidays over the subject field were

omitted from further consideration. Thus, only the best imagery

for each month of the corn growth cycle was selected for digi-

tization, registration, and multivariate analysis. This same

imagery was again reviewed to determine which photography had

the best contrast and clarity. For example, because of the

atmospheric scattering of blue light, the blue spectral image

in the multispectral photography proved to have little infor-

mation content. Consequently, the final data set selected

147

for the analysis consisted only of the green, red, and near-infrared spectral images of the selected frames of multispectral photography. Table 10 lists the multispectral photography that was selected for the digitization, registration, and multivariate analysis.

Site	Mission Number/Frame	Date	Bands Digitized	Scale
1	GR74-17C/Fm 13	1 May 74	B,G,R,IR	1:11000
1	GR74-39C/Fm 61	2 July 74	B,G,R,IR	1:12000
1	GR74-47C/Fm 43	16 July 74	B,G,R,IR	1:20000
1	GR74-60C/Fm 36	19 Aug 74	B,G,R,IR	1:20000
1	GR74-65C/Fm 56	4 Sep 74	B,G,R,IR	1:12000
1	GR74-86C/Fm 59	17 Oct 74	B,G,R,IR	1:17000
3	GR74-17C/Fm 40	1 May 74	G,R,IR	1:11000
3	GR74-39C/Fm 91	2 July 74	G,R,IR	1:12000
3	GR74-47C/Fm 49	16 July 74	G,R,IR	1:20000
3	GR74-60C/Fm 51	19 Aug 74	G,R,IR	1:20000
3	GR74-65C/Fm 75,76	4 Sep 74	G,R,IR	1:12000
3	GR74-69C/Fm 93	16 Sep 74	G,R,IR	1:20000
3	GR74-86C/Fm 26,27	17 Oct 74	G,R,IR	1:17000
7	GR74-14C/Fm 122	29 Apr 74	G,R,IR	1:16000
7	GR74-39C/Fm 157	2 July 74	G,R,IR	1:12000
7	GR74-52C/Fm 215	22 July 74	G,R,IR	1:50000
7	GR74-60C/Fm 98	19 Aug 74	G,R,IR	1:20000
7	GR74-69C/Fm 33	16 Sep 74	G,R,IR	1:20000
7	GR74-83C/Fm 65,71	8 Oct 74	G,R,IR	1:20000

Table 10: Table of multispectral photography digitized to support the feasibility study in crop yield prediction. Pertinent photographic data is listed for each agricultural test site; including mission number, frame number, date of mission, multispectral bands digitized, and original scale of the photography.

Table 10 illustrates some important points about the useable image data. With the six inch focal length of the multispectral camera lenses, one half of the photographic

148

scale of the photogrpahy indicates the altitude at which the multispectral photography was collected. The aerial photography collection program had been planned to acquire the medium altitude photography at 10000 feet above ground level (AGL). Unfortunately, due to cloud cover at various altitudes, exceptions had to be made which resulted in imagery being collected at 6,000 ft, 8,000 ft, and 8,500 ft AGL. Furthermore, Site Seven's data set indicates that one mission which was included in the digitized data set, was acquired at 25000 ft AGL (Scale: 1:50000) by the AFFTC U-2 aircraft. In the final analysis, by utilizing special registration software and digitization techniques, all of the imagery was digitized at approximately the same scale of 1:12000.

Though it did not matter whether the green, red, or near-infrared spectral image was digitized first (because the spectral images were at the same scale), some consideration had to be given as to which month's frame of multispectral photography was digitized first. Since the object of the digitization and registration was to produce a spectrally and temporally registered data set, a decision had to be made as to which month's imagery would be the standard scale to which the other frames of multispectral imagery would be registered to. In the case of Site One (scale range: 1:11000 to 1:20000), it was decided that the Site One image data set seemed less ambiguous if the small scale imagery was optically magnified

149

to the scale of the large scale imagery (multiplication of data) and then digitized.

The technique for adjusting the gross misregistration between imagery collected during different months was implemented in a simple fashion. The multispectral imagery of each site having the largest photographic scale was digitized first. For example, in the case of Site One, the green, red and near-infrared images from frame 13 of the May 1, 1974 multispectral coverage were digitized first. This frame's spectral images were displayed on a monitor prior to the actual digitization process. During that interval, the outline of the field site was traced with grease pencil on the glass face of the black & white monitor (Reference Figure 47, Page 152). After digitizing the three spectral images at the same focus settings of the Computer Eye Scanner, a new set of spectral images from another month's frame of multispectral photography was digitized. The scanner of the Computer Eye was adjusted in position above both the light table and the new frame of multispectral photography prior to each digitization. The scanner's position was adjusted until the outline of the small scale image of the corn field approximately matched the boundaries of the field that had been traced on the glass of the television monitor. When this approximation to gross scale was achieved, the three spectral images from the frame of multispectral photography were digitized without any further

adjustment to the optics of the scanner (because the three

spectral images were all photographed at the same scale).

Fine adjustments in misregistration were performed using the

DICIFER software developed for the registration of images with

different scales.

Digitization of the multispectral photography was per-

formed with the Spatial Data Systems (SDS) Computer Eye,

Digital Equipment Corporation (DEC) PDP-11/20 Minicomputer, and

the Richards Light Table. Figure 47 illustrates the Computer

Eye vidicon digitizer mounted vertically on a X-Y trans-

lational track situated over the light table. A Miratel

black & white monitor, at the left of Figure 47, permits the

user of the digitizing system to view and orient the image

being digitized prior to actual digitization. The black hood

surrounding the vidicon Computer Eye, Figure 48, prevents the

ambient glare of room lighting from affecting the digitized

imagery. The black film format masks were used to mask areas

of the photography that were not being digitized. Such masking

prevented the Computer Eye from being affected by the glare of

the light table.

After performing preliminary computer hardware setup

procedures, digitization was completed with the following steps:

1) clean both the light table glass and the multispectral photo-

graphy to be digitized; 2) position one spectral image of a

selected positive multispectral frame on the light table beneath

151
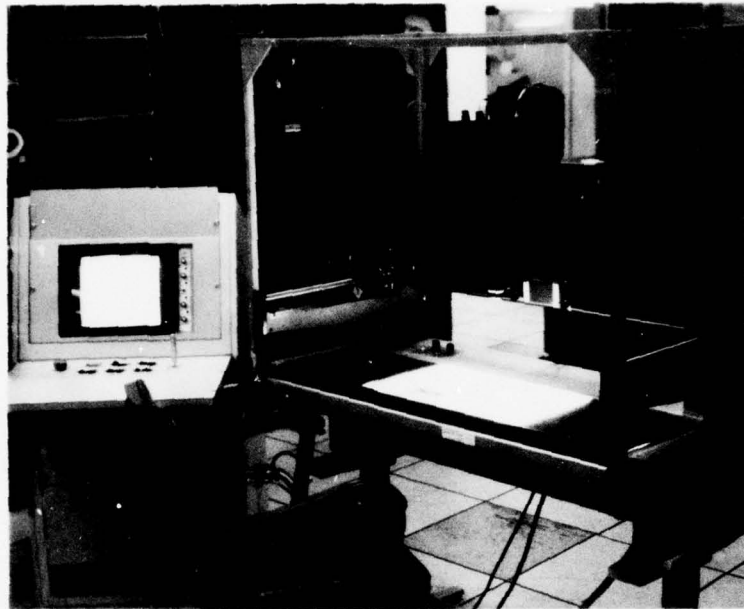
Figure 47. SDS Computer Eye's scanner situated on X-Y translational track above light table. Multispectral photography was digitized utilizing this system and the DICIFER software.

Figure 48. Digitization with the Computer Eye required a black hood and metal mask for excluding ambient room light and light table glare during digitization.

152

the Computer Eye scanner and orient the image while observing the black & white SDS monitor; 3) adjust the optical focus of the scanner until the desired scale and resolution is achieved; 4) adjust the contrast of the displayed image by adjusting the aperture setting on the scanner's lens system; 5) using the profiler capability of the Computer Eye, scan the entire image until the darkest area of the scene to be digitized is located and set the calibration control on the SDS monitor to zero; 6) repeat the process of step 5 for the brightest area of the entire scene; 7) using the interactive DICIFER software, call the computer program for digitizing with the Computer Eye system; 8) type in the name of the new digital image and execute the digitization process; 9) review the digital image on the television monitors using the display and statistics options. Accept or reject the digitized image.

The last step of the digitizing process required both a visual check of the digitized imagery, for such features as resolution and contrast and an analytical check of the actual grey levels of the digitized image. The latter check was accomplished by graphically creating an area file of representative portions of the new digital image and obtaining the statistics of the digitized image's grey levels contained within the boundaries of the area file. The statistics which are output appear as a histogram depicting the frequency of each grey level contained within the area file of the image.

153

If the picture is digitized improperly, saturation, or a skewing of the histogram to the low or high end of the horizontal scale will be observed. Though this skewness may sometimes be adjusted artificially using some of the DICIFER preprocessing algorithms (i.e., equal probability quantization, etc), a poor histogram was best adjusted by redigitizing the scene after checking and adjusting the calibration controls on the Computer Eye.

All images of the multispectral data were digitized such that the grey level values of the digitized image ranged from approximately 10 to 250. In some cases, certain spectral images of the photography were low in contrast due to the reflectance characteristics of the imaged materials for the particular bandpass and time of year (especially in the case of active vegetation with dynamic spectral signatures). Such pictures might only be digitized in a grey level range of 25 to 175. However, the important point is that the images were digitized such that the saturation or clipping did not occur in the final digitized scene. This technique, though unable to maintain the calibration or sensitometric control of the analogue film process, did limit the grey level values of the digitized images of the test sites to the linear region of the image density-exposure curve(Wenderoth and Yost,54).

Digitization Problem

The Computer Eye's scanner at the RADC Image Processing Facility contains a vidicon detection tube. This vidicon tube has a transmission response gamma of about 0.65. This means that if the Computer Eye were to digitize a photographic image into 256 grey levels, the digitized output value (Z) would be expressed as: $Z = KI^g$ /where "I" is the image brightness; "K" is the proportionality constant; and "g" is the transmission gamma, 0.65.

If the detector were a silicon diode type (available from SDS), gamma would approximate one, implying that the digitization process would be linear in digitizing the 256 grey levels. Unfortunately, with the vidicon tube, the Computer Eye digitizing process was nonlinear during this study.

To conclude meaningful results from data which had gross non-linear features due to digitization would be considered fallacious reasoning. Certainly both calibration and linearity during digitization would definitely enhance the multispectral imagery's utility for crop yield prediction. However, it is believed that the digitized imagery data set developed for this feasibility study is useable, despite its nonlinear characteristics, and any conclusions from the study can be relied upon. This belief is founded in the fact that the yield prediction model is based upon the analysis of temporal variations of the image density record of a crop's multispectral albedo (reference the yield prediction theory

155

section, pages 189-199). Since the nonlinear feature is characteristic of <u>all</u> of the digitized multispectral imagery for the months of April to October, any multivariate or regression analysis of the temporally-registered imagery would be effectively monitoring density variations of the complete nonlinear data set. The result of such analysis is that the nonlinear characteristics of the data are effectively normalized to a degree. Though the remaining information is probably slightly nonlinear, the greater portion of the data's information content may be attributed to the actual reflectance of the crop.

Introduction to Registration

Upon completing the digitization of each agricultural test site's set of multispectral photography, the digital data set was registered. Once registered, any digital image of any geographic location would be imaged in the same pixel (picture element) locations of each digital image of the set. The registration process was the key step in the development of a temporally-registered, digital multispectral imagery data set. Only if the image data sets were spectrally and temporally registered could the multivariate and regression analyses of pixel grey level data or groups of pixels' mean grey level data provide statistically meaningful results in the feasibility study of crop yield prediction.

Two techniques were implemented for registering the digitized multispectral images. The first technique compensated for coarse misregistrations between different months' multispectral imagery. This technique, performed during the digitization phase (reference digitization section), consisted of adjusting the optical focus of the Computer Eye scanner. By adjusting the scanner focus, small scale imagery was "optically enlarged" until the image viewed on the black & white monitor before digitization conformed to the boundary dimensions of a previously digitized large scale scene. The photographic scene

157

was then digitized. The resultant digital image was in close registration to the previously digitized large scale scene. The second registration technique provided for fine adjustments in registration and consisted of employing registration software developed for the RADC DICIFER Image Processing System. The registration algorithm, developed for RADC by contract and reported by Zanon(57), is described in the following section.

Image Registration Algorithm*

The registration routine, utilized during this study, automatically performed scale, translation, and rotation corrections based on a set of user-selected control points. The control points, determined visually by a user, represented the same ground position. For the sake of discussion, suppose that the intersection of two roads was represented by a picture element on each of two or more images, then these picture elements could be selected as "control points" on their respective images. As intuition suggests, one control point selected on each of the images will uniquely define a translation of one image so as to be "registered" with respect to the

---

*Note that this section describing the image registration algorithm was virtually copied from the RADC Technical Report entitled, Spectral Analysis (Zanon,57) with the verbal permission of the RADC Project Engineer.

reference image. The quotation marks around the word "registered" may be removed if, in fact, the translation defined by the pair of control points did actually apply to the whole image. If this is not the case, then it is possible that only the selected control points of the images would actually be registered upon the completion of the translation.

If it is known that translational registration will not suffice (or it is not certain whether it will suffice), two control points selected on each of two images will define a rotation of one image so that it will be registered with respect to the other. (It is possible that these points would, at the same time, define a translation with a scale change, but both of these effects would be absorbed by the algorithm that performs the rotation.) In the event that one image is a pure translation of the other, then accurate selection of the two pairs of control points would determine that no rotation is required. Similar to that discussed above, if one image is not a rotation (with scale change and translation) of the other, then it is possible that only the control points would actually be registered. This leads to the third type of misregistration that is corrected for by the registration routine implemented under this effort. The third option will handle not only the above cases, but also the case where there is a different scale

159

change in two independent directions (e.g., the scale factor in the row direction may be 1.0, while the scale factor in the column direction may be 1.2). Three pairs of control points suffice to uniquely define the parameters to perform the necessary registration to compensate for the above assumed mis-registration.

In the event that the user recognizes more than three ground features which may serve as control points, the extra points result in a possible loss to the uniqueness of the registration parameters. To permit the use of more control points, the registration routine automatically performs a least squares fit of the registration formula with the control points selected. This enables the program to compensate for possible observational error in the user's selection of the control points. When the routine displays the error encountered at each control point, the user can then judge the accuracy of the control point selection. We are of the opinion that it is better to have a list of errors inform the user that one of the eight control points is inconsistent with the others, than to use three control points to uniquely determine registration parameters that will not adequately register the images. The actual formula for minimizing the error using the least squares method is presented in Appendix C of this report.

Since it has been determined that it is sometimes difficult to find the same ground feature on all of the spectral images, the current image registration routine permits a different set of reference control points to be used for each image to be registered. This in no way degrades the resulting registration (again, emphasizing that the misregistration is assumed to be as described above). Various other options are available which increase efficiency and flexibility of the registration. These additional options are discussed in Appendix C.

Figure 49 illustrates how image II is registered with respect to image I to yield image III. Notice how some picture elements from II are lost in the registered image III while some picture elements of image III cannot be determined from image II.

Table 11 and Figure 49 illustrate the original misregistration of the control points selected on the respective images. Also included in the table and figure are the repositioned locations after a translation, followed by a rotation with independent scale corrections, was performed using the same set of control points.

In Figure 50, the arrows and point positions are drawn to the same scale, but the positions of the control points relative to each other are at a different scale than the arrows.

161
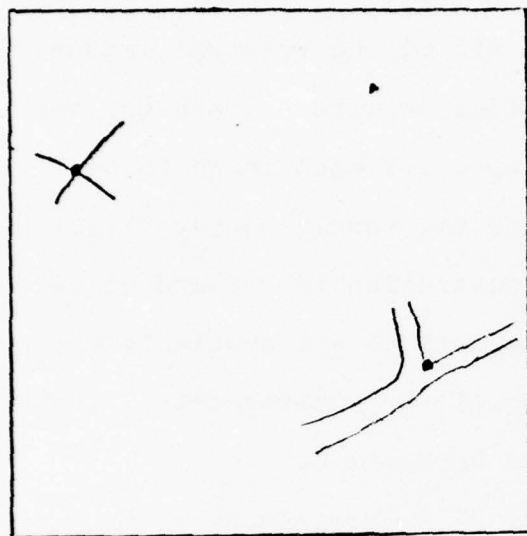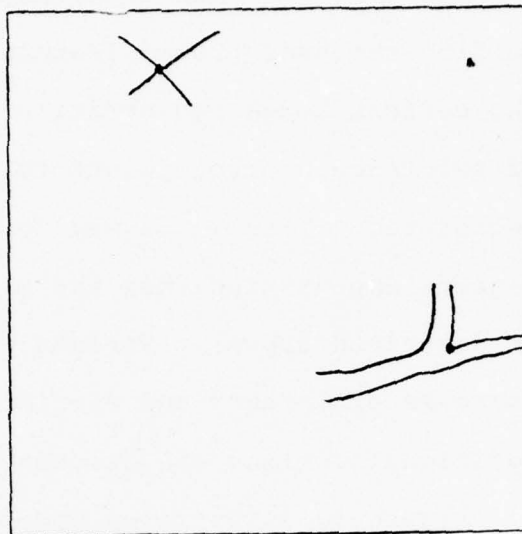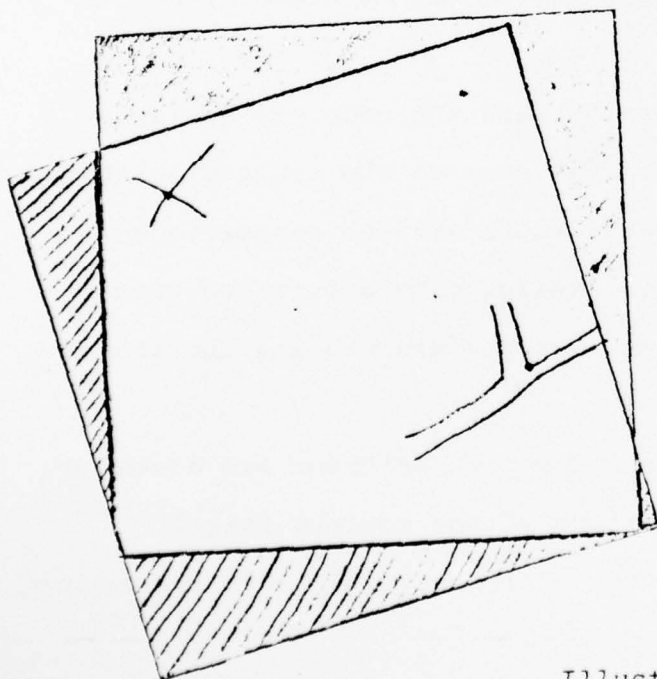
Image I

Image II

Image data not avail-
able from image II

Image data of image
II lost in image III

Figure **49.**

Image III

Illustration of the Effect of
Registration on 2 sets of
Registered Files

162

Control Point	Ref. Image Control Pts	Misregistered Image Control Pts	Original Error	Resultant Error (Translation)	Resultant Error (Rotation-Indep Scale)
1	445,541	445,542	Ø,-1	0,0	0,-1
2	961,699	961,702	Ø,-3	0,-2	-1,0
3	891,27	893,32	-2,-5	-1,-4	0,-1
4	326,127	328,127	-2,Ø	-1,0	0,0
5	50,206	54,205	-4,1	-3,1	-2,-1
6	139,548	138,546	1,2	1,2	1,0
7	321,784	320,783	1,1	2,1	0,-1

Table 11. Comparison of Control Point Errors
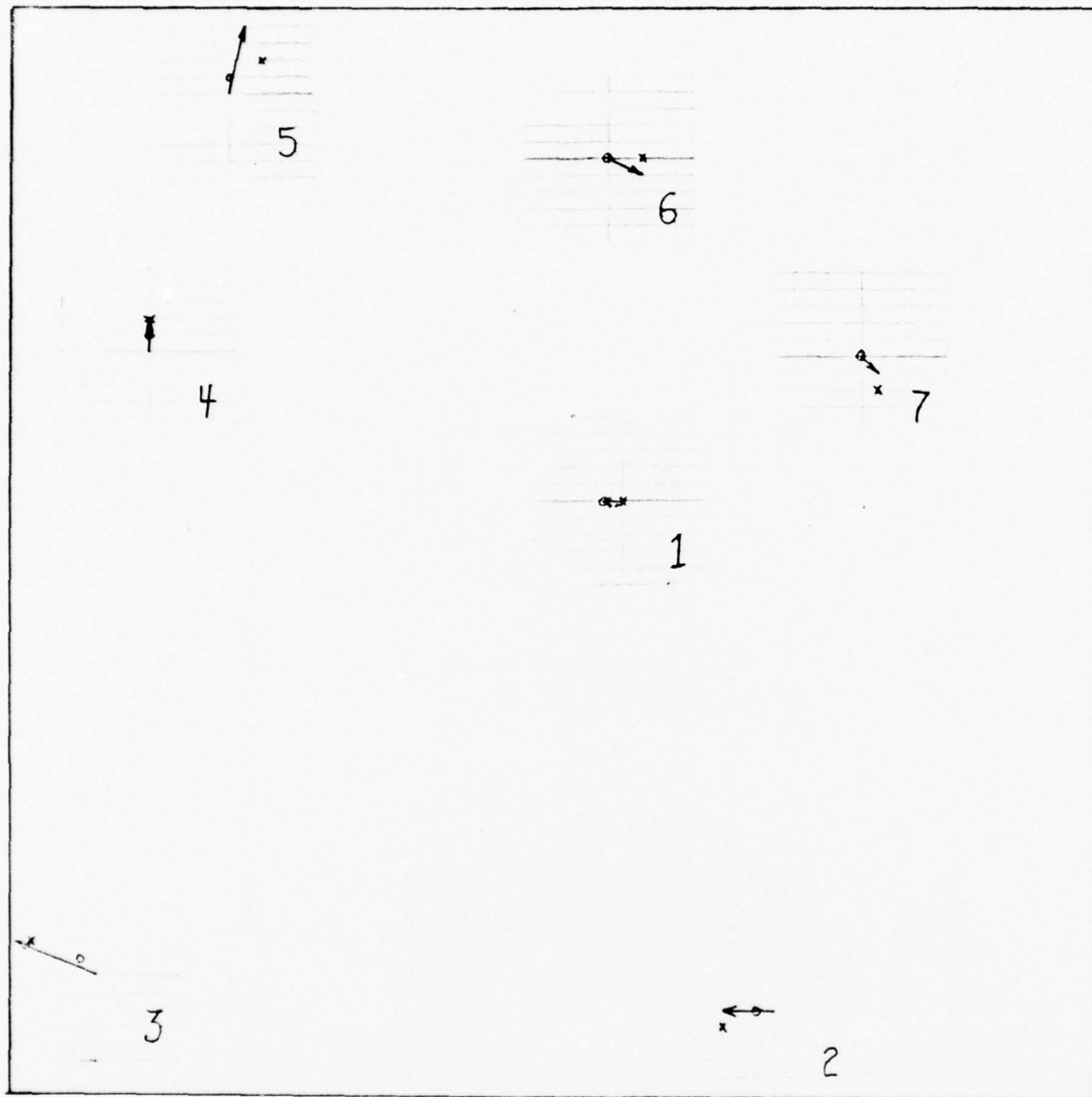
Figure 50.

Relative Control Positions with Magnified
Scale for point Relocation Positions

164

- The source of the arrow indicates the control point on the misregistered image.

- The arrow head indicates the control point on the reference image (and hence, the desired destination of control points).

- The symbol 'o' indicates the approximate relocated position of the control point assuming that only translational registration was performed, using all the control points.

- The symbol 'x' indicates the approximate relocated position of the control point assuming that rotational, translational, and scale misregistrations were corrected.

Registration Procedures

Despite the coarse registration techniques employed during the digitization to compensate for the different scales and misregistration of the multispectral photography, the registration software was required to fine tune the registration between the different months' multispectral digital imagery. To register two different months' digital images, the registration algorithm had to make small scale adjustments, rotations, and translations of pixels within the registered imagery.

A sequence of registrations was performed in order to produce the temporally-registered, multispectral imagery data set. Typically, the July 2nd multispectral digital images were registered first because of their large scale (1:12000). The green, red, and near-infrared images were all the same

165

scale (photographed simultaneously); consequently, only an initial optical adjustment of the Computer Eye's scanner was necessary. The near-infrared image was always the standard to which the green and red images were registered. This technique was utilized because the land marks or control points used for the registration and the boundaries of the corn field site were most distinct in the near-infrared spectral image. Upon completion of the registration of the July 2nd images, the digital images from other months' multispectral coverage were registered to the digitized and registered July 2nd imagery. This second phase of the registration was implemented by initially registering the near-infrared spectral image of an arbitrary month's digital imagery with the near-infrared image of the July 2nd imagery. Compared to the previous registrations of the July 2nd multispectral images which required only translation and rotation corrections, this second phase of registration required scaling corrections as well as translation and rotation adjustments. The green and red images of the second frame of digitized multispectral imagery were then registered to the newly registered near-infrared image. This last registration resulted in all of the two months' multispectral images being temporally and spectrally registered. This last phase of the registration procedure was repeated with the remaining months' multispectral images.

Though minor differences existed between the registration procedures for multispectral images of the same scale versus the registration procedures used for multispectral images with different scales, the basic procedure for registering images with the DICIFER registration software were as follows: 1) Display the digitized images on the black & white monitor. After referencing the original multispectral photography on the light table, locate at least four or five geographic points at the perimeter of the digitized scene that are easily recognizable in the two images that are to be registered (generally two images are registered at a time; i.e., near-infrared with green followed by near-infrared with red, etc.); 2) Using the interactive software program for image registration and the DICIFER display capability, the geographic points are selected as registration control points and stored in computer memory; 3) After the registration control points had been selected for the two images, the registration program computes a least squares fit of the registration formula, using the selected control points, and displays on a Tektronix CRT a measure of the error between the selected points (Errors of $\pm(1,1)$ pixel were tolerated during registration of the images); 4) Following an execute command by the user, the registration program computes the least squares best fit parameters for registering the two images and then produces a new image which is the second image

167
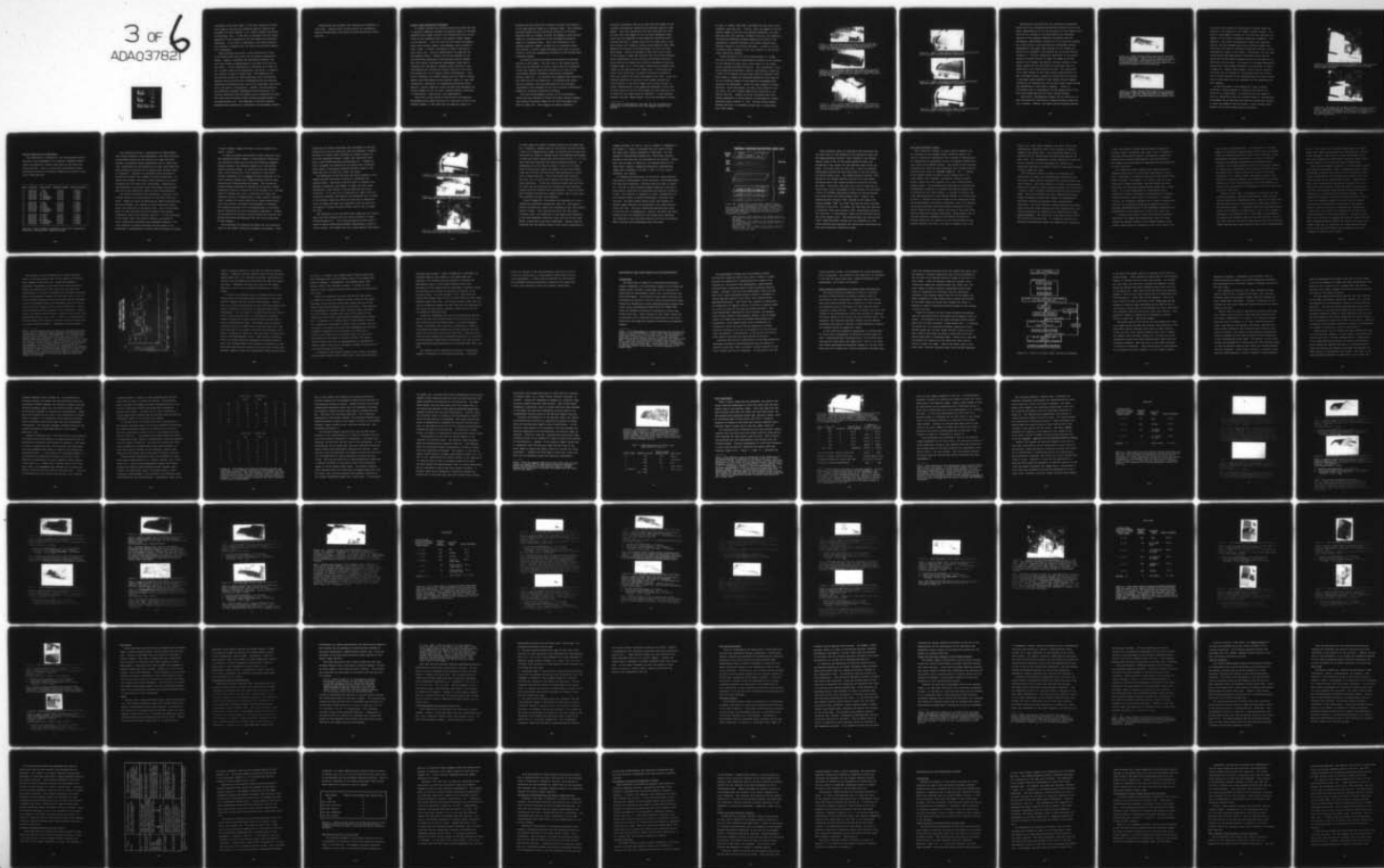
registered to the first image; 5) The user compares the registered image to the original reference image by checking the alignment of linear features (i.e., roads, property boundaries, intersections, etc). If the user is satisfied with the visual appearance of the registration, the two images are considered registered. If the user is unsatisfied, the entire registration process is repeated with the control points being reselected if necessary.

There certainly are errors in the registration of each of the agricultural test site's set of digital multispectral images. However, throughout the registration process, the error was limited to approximately a one pixel error per control point. If the registration is inaccurate by a pixel, at the (1:12000) scale of the photography, the error in registration amounts to about 20 square feet. The regression and multivariate studies for yield prediction could not possibly generate valid conclusions by investigating the multispectral reflectance fluctuations over time at the individual pixel level due to the error in registration. However, the multivariate and regression analyses investigated the multispectral reflectance variations over the growth cycle by analyzing the mean grey level data of several pixels occuping selected areas of the agricultural site. This approach in the data analysis generated both statistically significant and believable results.

Registration was extremely time consuming but essential in
this study. Computer time utilized during the registration
effort averaged about forty hours per each agricultural site's
data set.

Digital Image Combination Procedures

An unusual problem that occurred during this study was that of digitally combining together two spectral images of the same bandpass which imaged one-half of an agricultural site in one frame and the remaining half in the adjacent frame (camera cycling problem). Since the film for each frame was exposed every three seconds (camera intervelometer rate at 10000 ft (AGL), Scale - 1:20000), the amount of solar illumination, solar angle, etc., would be approximately the same for the two adjacent frames. However, due to the problems of the man-controlled production of the balanced positive imagery from the original multispectral photography, there was no guarantee that the density contrast within one-half of the photographed site matched the density contrast of the other half-imaged site in an adjacent frame of photography. Site Three's September and October imagery and Site Seven's October imagery were handicapped by this problem (Table 10, Page 148). Consequently, production of a temporally-registered, multi-spectral, digital image set, which included this September and October imagery of the two sites, required special techniques in digitization, registration, and preprocessing.

During the initial review of the positive multispectral photography which imaged one-half of a test site in each of two adjacent frames, it was noted that any spectral image of a

170

balanced positive frame had a different density when compared to the same spectral image of an adjacent frame. This phenomena occurred because the man-controlled production of balanced positives does not attempt to match the gammas of each spectral image of a balanced positive to the corresponding spectral image of an adjacent frame. Since this difference in two adjacent spectral images' contrast was not corrected during digitization, a digital gamma adjustment would have to be performed by using the preprocessing software of the DICIFER Image Processing System.

The typical digitizing process was followed as described earlier in this report. The left half of the imaged agricultural site was adjusted in scale to match the site boundaries of the July imagery which were sketched on the face of the digitization monitor (reference digitization procedures section, page 147). The procedure was repeated when digitizing the right half of the imaged site. By following the registration procedure, the digitized halves of the site were registered to the remainder of the site's digital multispectral imagery by using the registration software.

Since the intervelometer setting on the multispectral camera had been adjusted to obtain 60% stereo overlap coverage, the resultant digitized images of the field overlapped each other by about 60%. This overlap was helpful because it

171

occupied a geographic area of the site that was imaged on two
adjacent multispectral frames having different spectral image
gammas. After the registration had been performed such that
the left and right images of the site were separately regis-
tered with the remainder of the temporally-registered multi-
spectral data set, a graphic area file was drawn using a joy-
stick cursor and a computer display which permitted grey level
sampling of the area of overlap between the left and right
half-images. Density histograms of each area file's pixels
of each spectral image were obtained with the DICIFER software.
After consideration of other factors which might determine
which half of the imaged site would be considered to have the
standard gamma (i.e., cloud cover, percentage of useable field
imaged on the particular half frame, etc), a half-image was
chosen which would have its gamma artificially adjusted to
match the gamma of the other half-image of the site. Since the
original densities of the site's half images occupied the
linear portion of the film's gamma curve (D-log E curve), a
linear transformation of the gamma was performed to bring the
contrast levels of the two half-images to match along the line
at which the two images would be combined.* A user computer
program, entitled "Range Change", which was developed in-house

_____

*Note that no consideration was made for any nonlinearities
induced by the digitization with the <u>Vidicon</u> Computer Eye
Scanner.

by Capt. W. Rogers, RADC/IRRE, performed the grey level trans-
formation task very well. Finally, after the gammas of the two
digital images of the site were matched adequately, the soft-
ware for area file creation, Hadamard filtering, and computing
the absolute difference between images was utilized in the
generation of complete September and October digital multi-
spectral images of Sites Three and Seven. Figures 51,52,&53
illustrate three examples of the final product of the digital
image combination process.

Partly because of its size of approximately 2.5 acres,
Site One's multispectral photographic coverage was not degraded
by the problems of cloud cover, pilot error, or the camera
cycling problem described previously. Upon completion of Site
One's temporally-registered, digital, multispectral image set,
a total of 24 computer pictures were stored on magnetic disks.
This number of images was generated because all four bands of
the six different frames of multispectral photography were
digitized and registered. During the later studies in stress
detection, yield measurement, and crop yield prediction only
the green, red, and infrared images were considered in the
digital data set. Figures 54,55,&56 illustrate Site One's
digital appearance in three channels of multispectral imagery
collected during October 17, 1974. Another fifteen images
spanning the April to September period were in registration
with these images.

Figure 51. Near-infrared spectral image of Site Three after digitally combining and preprocessing the two halves of the imaged site recorded on two adjacent digitized frames of October 17th multispectral photography.

Figure 52. Green spectral image of Site Three after digitally combining and preprocessing the two halves of the imaged site recorded on two adjacent digitized frames of October 17th multi-spectral photography.

Figure 53. Near-infrared image of Site Seven after digitally combining and preprocessing the two halves of the imaged site recorded on two adjacent digitized frames of October 8th multi-spectral photography.

174

Figure 54.   PSAF22, Green spectral image of Site One digitized from the October 17th multispectral photography.

Figure 55.   PSAF23, Red spectral image of Site One digitized from the October 17th multispectral photography.

Figure 56.   PSAF24, Near-infrared image of Site One digitized from the October 17th multispectral photography.

175

Agricultural Site Three was only partially photographed during one of the selected multispectral missions due to pilot error; approximately 10% of the top portion of the imaged field could not be included in the multivariate and regression analysis of the complete temporally-registered data set. Furthermore, approximately another 8% of the site was shadowed by a cloud during a July multispectral photography mission. Consequently, the upper right quadrant of the imaged site could not be included in the analysis of the entire data set (Figure 57). Finally, because of limitations in the preprocessing software's ability to match the gamma of the two halves of the October multispectral digital imagery of Site Three, another 10% of Site Three was eliminated from consideration in the Yield Prediction Study. The limitations of an older version of the range change algorithm(Lietz,32) had introduced density clipping or saturation of the pixels in the upper right quadrant of the October digital imagery (algorithm was replaced with the in-house range change program and implemented on Site Seven's imagery). Figure 58 illustrates what is considered to be the useable area of the Site Three image data set for April through October.

Site Seven's multispectral imagery did not suffer all of the difficulties of Site Three's imagery because clouds were not a problem. However, the camera cycling problem combined

176
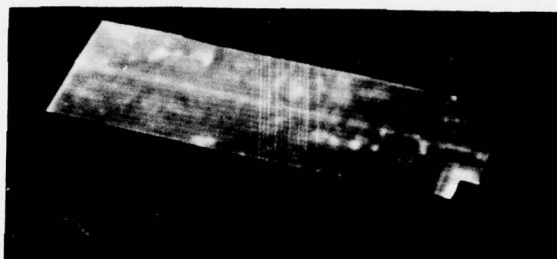
Figure 57. DIFFØ2, Edited digital image of the digitized near-infrared spectral image of the 17 October 1974 multi-spectral photography. Pilot error and cloud cover resulted in the top and upper-right hand portions of the entire registered data set of Site Three to be eliminated from the multivariate analysis for yield prediction.

Figure 58. DIFFØ3, Edited digital image of the digitized near-infrared spectral image of October 17th multispectral photography. This area represents the only useable pixel data of Site Three's complete temporally-registered, multi-spectral imagery.

with pilot error forced the techniques of digital image combination to be employed on Site Seven's October imagery. The error in photographic coverage of Site Seven was large and the imagery of Site Seven in adjacent frames was inadequate. Consequently, two October multispectral frames which were not adjacent and were photographed during different passes over Site Seven were used for obtaining relatively complete, digital multispectral images of Site Seven. Fortunately, the sun angle relationships were approximately equivalent during the two passes and minimal gamma adjustment had to be performed on the two halves of Site Seven. The results of the image combination process and digital gamma adjustment is illustrated in Figure 53, page 174. Figures 59&60 present various digital scenes of Site Seven for the purpose of illustrating both the location of Site Seven relative to the background and the actual useable area of Site Seven after all problems with the image data set were remedied.

At the conclusion of the feasibility study, hindsite indicated a better approach to digitally matching the two halves of an imaged scene. By digitizing the two images of the site sequentially from the separate frames of multispectral photography and by adjusting the digitizer calibration controls to match the gammas of the two scenes, a more reliable match between the two half-images could be obtained.

FIELD 7 W/RESP TO BACKGRND
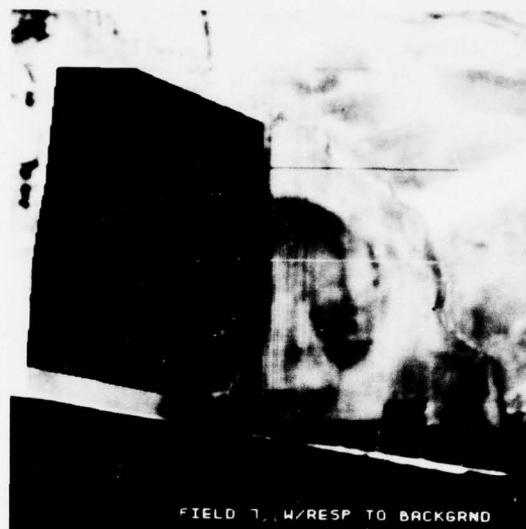
Figure 59. The black area of this September 16th near-infrared digital image illustrates the boundaries of the Site Seven agricultural test site.
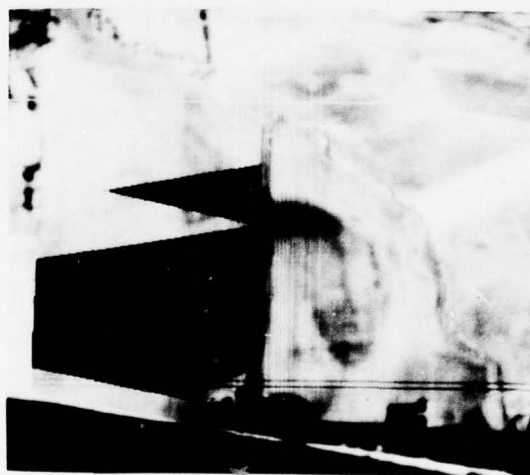
Figure 60. The black area of this September 16th near-infrared digital image illustrates the location of the useable pixel data in Site Seven's temporally-registered, multispectral data set spanning the corn growth cycle of April to October in Central Pennsylvania.

Digital Image Data Set Description

The digitization, registration, and preprocessing efforts resulted in the development of a complete, temporally-registered, multispectral, digital image data set for Sites One, Three and Seven. Table 12 lists which multispectral aerial photography missions and spectral images were included in each site's image data set.

Site	Mission#	Date (1974)	Spectral Images	Effective Scale
1	GR74-17C	1 May	G,R,IR	1:12000
1	GR74-39C	2 July	G,R,IR	1:12000
1	GR74-47C	16 July	G,R,IR	1:12000
1	GR74-60C	19 August	G,R,IR	1:12000
1	GR74-65C	4 September	G,R,IR	1:12000
1	GR74-86C	17 October	G,R,IR	1:12000
3	GR74-17C	1 May	G,R,IR	1:12000
3	GR74-39C	2 July	G,R,IR	1:12000
3	GR74-47C	16 July	G,R,IR	1:12000
3	GR74-60C	19 August	G,R,IR	1:12000
3	GR74-65C	4 September	G,R,IR	1:12000
3	GR74-86C	17 October	G,R,IR	1:12000
7	GR74-14C	29 April	G,R,IR	1:12000
7	GR74-39C	2 July	G,R,IR	1:12000
7	GR74-52C	22 July	G,R,IR	1:12000
7	GR74-60C	19 August	G,R,IR	1:12000
7	GR74-69C	16 September	G,R,IR	1:12000
7	GS74-83C	8 October	G,R,IR	1:12000

Table 12. List of imagery included in each site's temporally-registered, multispectral imagery data set.

This feasibility study's investigation of semiautomatic crop stress detection, yield measurement, and yield prediction was performed using only the three site's image data sets. However, an important study consideration was to insure that this investigation of the feasibility of crop yield prediction would generate conclusions at a statistically reliable level of confidence. Certainly no confidence could be placed on conclusions derived from a sample of three observations. To remedy this dilemma, the ground truth information and image-extracted data of sixteen corn "minifields", located within the boundaries of the three large sites became the principal data set of the feasibility study of crop yield prediction.

The sixteen minifields were selected as a function of the stresses acting on the crop planted at Sites One, Three and Seven. Though the yield of approximately forty minifields were actually measured, only sixteen of these sites were located within the boundaries of the three sites. Each of the sixteen areas had been monitored during the ground truth program (reference: ground truth section, page 65). At the conclusion of the growth cycle, but prior to harvest, the yields of these 1/500th acre minifields were accurately measured. The final selection of these minifields had been based on the photographic interpretation of image truth photography collected

181

in early October (Kodak 2424 black & white infrared film,
scale - 1:3000).

In order to perform semiautomatic stress detection using
the registered digital images, a classification "design set"
was created for each site by using the television displays,
display graphics, and software capability of the DICIFER
Image Processing System. By extracting the tone features
from a site's design set, which consisted of the pixels'
density information of the imaged minifield locations of the
site's spectral imagery set, imaged stresses could be semi-
automatically differentiated and mapped. The interactive
process briefly consisted of displaying the spectral images;
graphically creating a design set of the classes (i.e., various
types of stressed corn, etc.); evaluating the image tone
features' ability to differentiate a class; select a decision
logic; classify the entire test site or image automatically;
check the results, and modify the procedures if necessary.
This mapping procedure will be elaborated upon in the report
section describing the yield measurement experiments. However,
the procedures for creating a digital design set from the tem-
porally-registered multispectral data set will be explained
in this section.

The procedures for creating the design set are relatively
simple if the proper interactive software is available. After

identifying the pixel coordinates that correspond to the mini-
fields within the site where the yield was measured, a graphic
box area is created, which surrounds the subject pixels, by
using the television display, cursor, and interactive soft-
ware of the DICIFER Box Area routine(Lietz,32). Figures 61,
62&63 illustrate the locations of the design set's minifield
areas imaged in arbitrarily selected spectral images of the
image data sets of Sites One, Three, and Seven.

The dimensions of each box area file which bounded a site's
imaged minifield contained 72 pixels of information per
spectral image. The actual area file which is stored in the
computer's peripheral disk memory is simply the four image
coordinates of the corners of the box defining the site's
minifields. If an area file is specified by "name" using the
interactive software, the computer can manipulate the 72 pixels
of the spectral image with which the area file is associated.
Since the image data sets are registered, each box area file
can specify 72 pixels contained within each of the 18 images
per site.

The dimensions of the box areas which comprised this study's
design set were selected with the "Foley Criteria" of image
processing(Foley,17). This criteria basically states that in
order to obtain statistically optimal semiautomatic classifi-
cation results, the minimal size of an image-sampled area should

Figure 61. Near-infrared image of Site One with black area files defining position of the site's minifields.
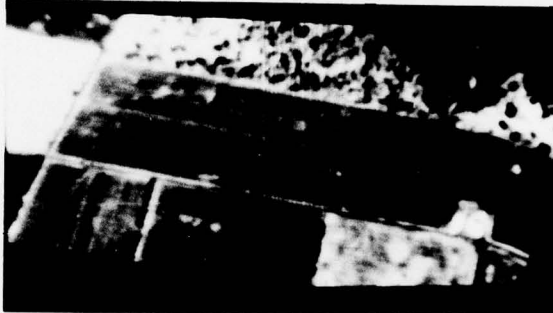
Figure 62. Near-infrared image of Site Three with black area files defining position of the site's minifields.
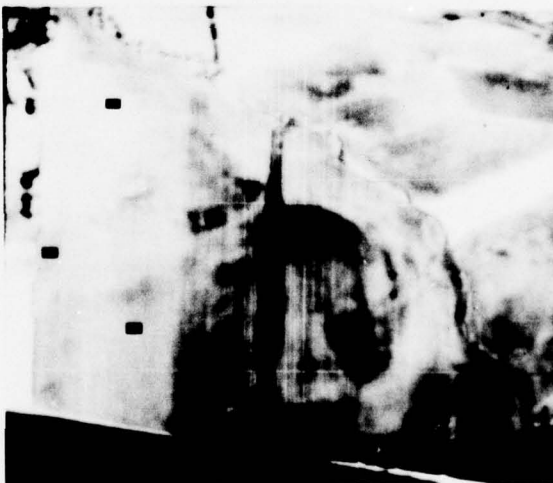
Figure 63. Near-infrared image of Site Seven with black area files defining position of the site's minifields.

184

be three times the number of measures making up the image data set. Therefore, assuming that all four bands had been utilized, a total of 24 measures ((the number is derived from the fact that there were six dates of useable aerial multispectral photography (4 bands per frame) during each site's growth cycle, 6 x 4 = 24)) would have been available for image processing analysis. By implementation of the Foley Criteria, a total of 72 pixels would be the minimal design set sample size for each stress class that would be differentiated semiautomatically in each site's imagery data set. Since three bands were actually utilized, only 54 pixels were necessary for each area file's pixel population. It was only determined after the area files for Sites One and Three had been created that the blue band imagery would be rejected from consideration. Since 54 pixels were minimally required for optimal results, the 72 pixel area files proved to be more than adequate.

Figure 64, page 187, illustrates the structure of a site's temporally-registered multispectral data set. Each month's digital multispectral images are registered to the other months' spectral images. This registration means that any pixels contained within the boundaries of the imaged site correspond to the same geographic positions within the site as the other images' pixels having the same coordinates. Though Figure 64 indicates that the specific dates of the useable multispectral

185

imagery are April 29, July 2, July 16, August 19, September 4, and October 17, Table 12 indicated that for certain sites, the dates were slightly different in some cases. For the purpose of demonstrating feasibility, these minor inconsistencies in the data were not considered to be critical. Therefore, for the sake of simplicity, the general dates of the three sites' useable, temporally-registered, multispectral images were considered to be May, 2 July, 16 July, August, September, and October.

To complete the description of the actual image-extracted data used for semiautomatic stress detection, yield measurement, and crop yield prediction, reference should be made to Figure 64. Each set of three colored rhomboid patterns illustrate a given month's digitized and registered multispectral images of a site. This pattern is repeated six times, corresponding to six different dates of useable photographic collection over the site. The piano shaped figure within each rhomboid or digital image represents the boundaries of an imaged agricultural test site. The pixel coordinates defining each image of the site are in registration. Finally, the small box areas located within the boundaries of the imaged site, represent the locations of the area files defining the site's minifields where accurate yield measurements were collected.

186

TEMPORALLY REGISTERED MULTISPECTRAL DIGITAL DATA
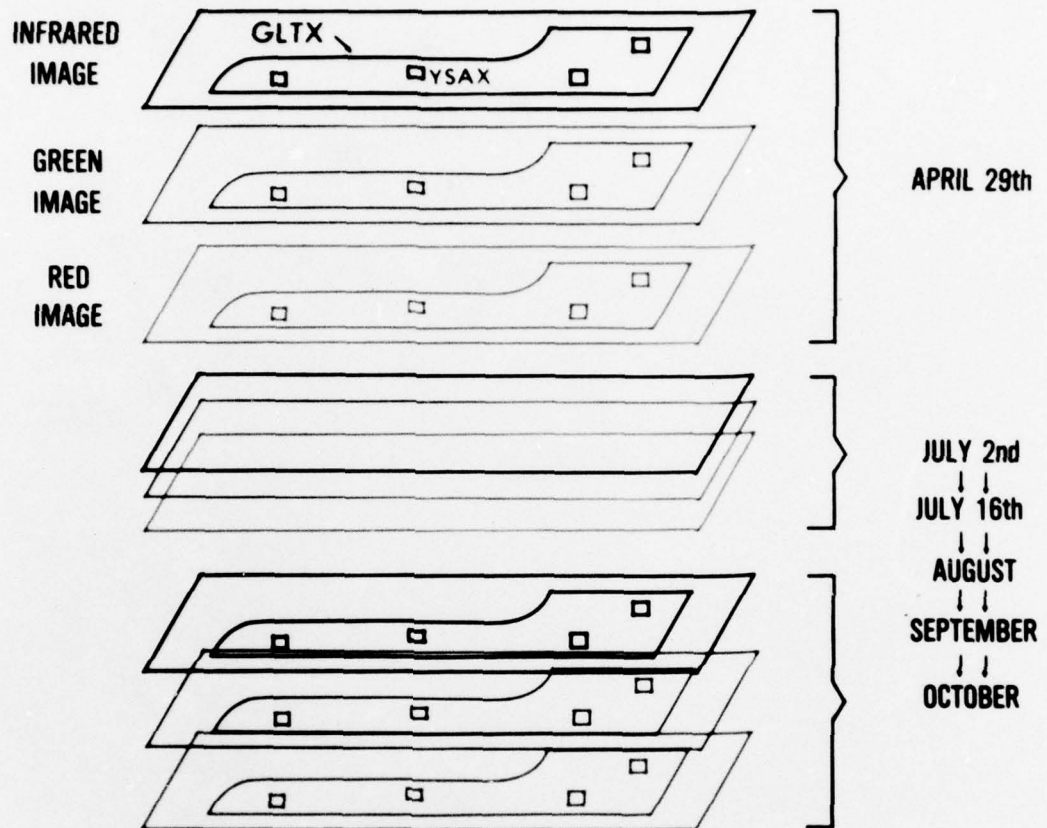
Figure **64**: Illustration of the digital, temporally-regi-
                stered, multispectral image data set assembled
and utilized in the semiautomatic crop identification, stress
differentiation, and yield prediction studies. The image-
extracted variables used in both the multivariate and regres-
sion analyses' studies may be more fully understood upon
referencing this figure:

- Image variable "GLTX" equated to the average density of
  all pixels contained within the site boundaries per band
  per month.
- Image variable "YSAX" equated to the average density of
  all pixels contained within each minifield's area file
  per band per month.
- Image variable "DIFX" equated to the numeric difference
  between "YSAX" and "GLTX" for each minifield's area
  file per band per month.

Three different types of information were extracted from the grey level or tone features of the registered image sets. The image-extracted variable "GLTX" equated to the average density value of all of the pixels contained within the boundaries of the imaged site per spectral image (band) per month. This meant that for each site, a total of eighteen measurements defined the mean grey level of the site during the entire growth cycle. The image-extracted variable "YSAX" equated to the average density value of all of the pixels contained within each imaged minifield's area file per band per month. This meant that each area file (16 total for all three sites) which corresponded to the location of a minifield, a total of eighteen measurements defined the minifields mean grey level during the entire growth cycle. Finally, the image-extracted variable "DIFX" equated to the numeric difference between "YSAX" and "GLTX" per band per month for each minifield's area file. As before, for each area file defining a minifield, there were eighteen "DIFX" measurements defining the "variance" in multispectral albedo of the minifield during the entire growth cycle. The rationalization for defining the "DIFX" image-extracted variable will be explained in a following section which describes the theory which structured this crop yield prediction feasibility study.

Crop Yield Prediction Theory

When scientific interest in remote sensing expanded with the launching of the LANDSAT-I(ERTS) satellite system, the trend in scientific investigation was directed to demonstrating the feasibility of estimating national or regional agricultural crop yields using LANDSAT imagery. The review of the literature indicates that there are many techniques being investigated for predicting the yield of crops(**Ref: pages 25 - 39**). Several of the models require extensive ground-sampled data (i.e., plant height, soil moisture, row spacing, number of barren plants, stalks per meter squared, etc.) and a sophisticated sampling model. The investigations dedicated to achieving the technological capability to predict crop yields remotely may be categorized into three phases: 1) attempt to identify the crops semiautomatically using LANDSAT imagery as a prime source of data; 2) estimate the total acreage of the identified crops using semiautomatic mensuration algorithms; and 3) by combining the acreage estimates with known weather parameters and per- tinent ground truth data, predict the yield of the crops using various regression analysis techniques. Unfortunately, due to the wide detection bandwidth of the LANDSAT multispectral scanner system and the inability of the investigators to tem- porally-register more than a few sets of imagery, only a few

189

crops (i.e., rice, wheat, soybeans, and corn) can be semi-automatically identified with a fair accuracy at certain stages of the growth cycle. High accuracies appear to be achieved only in regions where the particular crop is the major crop. As a consequence, acreage estimates of crops are not too reliable and yield estimates are forced to be inaccurate (reference literature review of crop identification studies, pages: 340- 353).

The literature review of studies investigating the feasibility of remote crop identification identified only two proven remote sensing techniques for accurate remote crop identification. The first technique utilizes a multispectral sensor with selected narrow-bandpass detectors for highlighting a specific crop's spectral reflectance differences from other crops. The second approach utilizes temporally-registered frames of LANDSAT multispectral imagery for demonstrating relatively accurate semiautomatic crop identification capability. The yield prediction model developed and investigated during this study employs both crop identification techniques innovatively to demonstrate the feasibility of crop yield prediction. The variations in the applications of the two original techniques were: 1) instead of utilizing the narrowband multispectral camera for identifying or speciating

190

a crop, the system's filters were dedicated to detecting
stresses within a particular crop (corn); and 2) instead of
analyzing temporally-registered, multispectral imagery for
demonstrating accurate semiautomatic crop identification,
the concept was investigated for demonstrating the feasibility
of yield prediction.

The literature review indicates that physiologically-
stressed areas within an agricultural crop achieve a yield
that is positively or negatively affected by the stress acting
within the area. For example, if a certain area of a corn
field had a mineral deficiency, the corn within that area
would have an inefficient photosynthetic process and a lower
yield. It is also established by investigation that a
vegetation's dynamic reflectance characteristics or spectral
signature becomes altered when the vegetation is stressed.
Finally, the literature indicates that a crop's photosynthetic
potential or yield may be measured by analysis of crop's
recorded albedo (reflectance of the vegetative canopy)which is
proportional to the canopy mass, leaf to soil ratio, etc.

If stress levels within a corn field could be detected,
identified, and mapped semiautomatically within an image,
perhaps the temporal history of the imaged stresses' pixel
density levels might be correlated to the final yield of the

191

crop within the stressed areas. In other words, if a photo-
graphic or electronic imaging system could be used to speciate
crops and differentiate stress levels within the crop, it may
be possible to measure and predict the yield of a crop by
analyzing the temporal fluctuation of the crop's albedo
recorded in the imaging sensor's temporally-registered imagery.
Of course, such a prediction, which might be feasible early in
the growth cycle, would be valid only if acts of God were
discounted (i.e., floods, fires, catastrophic events, etc).
Though the literature supports the feasibility of this theory
(Colwell,11), until this feasibility study was performed, no
research had investigated or demonstrated crop yield prediction
using real image data.

Figure 64, page 187 , illustrated the nature of the tem-
porally-registered, narrowband multispectral image data set
analyzed during this feasibility study. Each of the three large
agricultural sites had digital image sets that were created in
the following manner: 1) excellent quality multispectral photo-
graphy was collected on or about May 1, July 2 & 16, August 19,
September 4, and October 17th; 2) each frame of multispectral
photography was processed such that the four spectral images of
the given site had the same minimum density and range of density
levels spanning the linear portion of the D-log E characteristic

192

curve; 3) the green, red, and near-infrared spectral images within each frame were digitized and registered such that the actual geographic locations at the site were imaged at the same pixel coordinate in each month's spectral images (registration accuracy was +1 pixel or about 20 square feet).

There are two degrading features of the registered imagery that should be explained. First, precautionary methods were implemented to develop the film such that the density levels within the boundaries of the imaged site occurred within the linear portion of the characteristic curve. Unfortunately, there is no guarantee that the density minimum and density range of the four spectral images are equal in the balanced-positive, multispectral frame because the balanced-positive production is a subjective man-controlled process. The objective of matching the minimum density and density range of the four spectral images is to insure that the images display only the reflectance differences of the subject material (which has different characteristic reflectances in each band) as a function of bandpass and not due to the variable exposure in each image(Yost,54). Secondly, though great care was taken to digitize the spectral images such that their respective minimum density and density range were preserved, there was no guarantee that these objectives were met by using a man-controlled vidicon digitizer (the non-linearity of which is well known).

193

The effects of these aforementioned factors manifests itself in the mean density level of the pixels of the registered imagery of the corn site. Figure 65* presents a conceptual illustration of the basic crop yield prediction theory and image data set that this study was based upon. Interpretation of Figure 65 indicates that for May's digitized frame of multispectral images, the mean grey level of all of the corn pixels contained within the imaged site are brightest in the near-infrared image, with decreasing levels of brightness in the blue, green, and red images respectively. Due to the previously mentioned problems of film processing, digitization, and density range matching (reference the Digital Image Combinations section, page 170), there is no promise that a spectral image's mean density is actually greater or less than an adjacent spectral image's. Therefore, the mean density

---

*Note: The horizontal scale of Figure 65 indicates the months during which an imaginary ideal multispectral photography set was collected, digitized, and registered. The vertical scale reflects the mean digital grey level value of all pixels contained within the boundaries of an imaged field of corn (0-darkest level, 255-brightest). The mean grey level of the imaged corn field is plotted as a horizontal line which is color-coded as a function of the spectral image from which the mean value was computed. The mean grey level of pixels contained in stressed areas of the corn field are illustrated as a "variance hump" above or below the mean density value of the imaged field (horizontal line). Keep in mind that Figure 66 is only a conceptualization of what could occur but does not necessarily reflect the exact density fluctuations observed during the course of this study.

194

**1974 CORN GROWTH CYCLE
CENTRAL PENNSYLVANIA CORN**

255

MEAN IMAGERY DENSITY LEVEL
OF CORN FIELD × (PER BAND)

MAY JUN JUL AUG SEP OCT NOV
1974
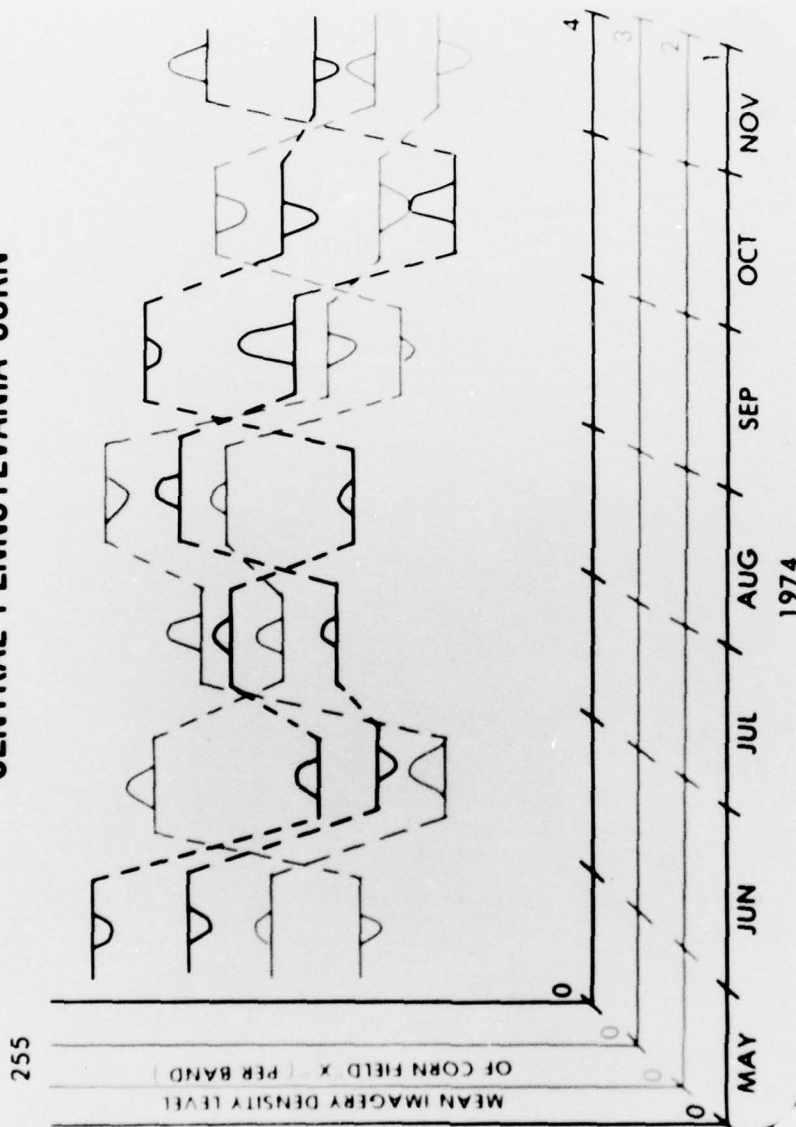
Figure 65: Concept illustration of the yield prediction model investigated during this feasibility study. The study utilized narrowband multispectral imagery which was digitized and temporally registered. The photographic information that was common to the imaged scene over the entire growth cycle was the noise level of the film. Since the film was processed such that the mean density level of imaged subject material, corn, occupied the linear region of the film density/exposure curve, the health of small areas with the corn field would be characterized by the variance in density that was greater or less than the mean density level. It was the analysis of the multispectral fluctuation of density variance over time that established a significant relationship with the yield of the crop.

195

level is actually greater or less than an adjacent spectral image's. Therefore, the mean density levels of the digitized image depends upon film processing controls, digitizing procedures, and the exposure of the film at the time the mission was flown. Therefore, the mean grey level of the imaged material or site is not an absolute reference of a crop's albedo.

If there is no absolute scale or reference for the mean density level of the imaged field, the utility of the multi-spectral imagery data set may seem questionable. However, there is a saving characteristic of the registered imagery that makes it useable for this study of _relative_ image density differences. This attribute is the fact that the mean density level of the digitized image, if translated to intensity values, occupies the linear portion of the original film's D-log E characteristic curve. This means that any density variations within an imaged area of the corn field site, for which the mean of the pixels' densities was determined (GLTX region & variable of Figure 64), also occupies the linear portion of the characteristic curve. These density variations within the imaged field can correspond to stressed areas of corn. The stressed areas can be mapped and identified in the digitized image data set because of the abilities of the multi-spectral camera's filters to differentiate stress levels within

196

the field. Of course, the stressed areas' density levels are also considered when the mean density level of the imaged corn field is computed. Consequently, the stressed areas' mean density levels (YSAX variable's region in Figure 64) constitute a linear variance from the mean density of the digitized corn field.

Figure 65 illustrates the fact that for any area within an imaged corn field, the average density levels of the pixels that make up the digitized area have a value that may be different from the mean density level of the field. This variance is illustrated as "variance humps" which are above (brighter) or below (darker) the mean grey level of the pixels of the imaged corn field (GLTX image-extracted variable value, Figure 64). The amount of a stressed area's imaged density variance is a function of the degree of stress, type of stress, and the interval in the growth cycle when the photography was collected. Therefore, regardless of the location of a pixel with a specific grey level within the imaged field, as long as the mean density of the field is linearly located, the density variance within the field is characteristic of a positive or negative stress which may be ultimately related to the yield of that specific area of the site.

To highlight these imaged stressed areas' density variances from the mean density level of the field, three image-extracted

197

variables were created. Figure 64,page 187, illustrates the variables meaning with respect to the imaged data set. Basically, the variable GLTX was created to identify the mean density value of the pixels contained within the boundaries of the imaged site for each month's spectral images. Two variables were created to describe the density of the stressed areas. The first variable, YSAX, identified the actual mean density level of the 72 pixels making up each imaged stressed area's area file. The second variable, DIFX, was the numeric difference between the YSAX value and the corresponding GLTX value for each month's spectral images of the site and its associated stressed areas.

Figure 65 illustrates that the mean multispectral density values for both the stressed areas of the field site and the site fluctuate over the growth cycle. It is this relative density fluctuation over time that is theorized to relate to the progressive health condition and final yield of the crop within the particular area of the field site. This study will determine the feasibility of predicting crop yield by analyzing the multispectral image density fluctuation over time by using image-extracted data described by the variables YSAX, DIFX, and GLTX.

The analysis of the temporally-registered, multispectral imagery is presented in the following sections. The first

198

section is devoted to the image processing techniques utilized during the investigation in semiautomatic stress detection and yield measurement. A later section describes the multivariate and regression analyses performed to determine the feasibility of crop yield prediction using only remotely sensed data.

Semiautomatic Crop Stress Mapping and Yield Measurement*

Introduction

The three sets of temporally-registered multispectral images, developed in the preliminary stages of this image processing study, provide a unique data set for studying the feasibility of semiautomatic stress level discrimination and yield measurement. This section of the report describes the procedures and results of the two related studies which demonstrated the feasibility of both semiautomatically mapping crop stress levels and measuring yield within an imaged crop, using the temporally-registered multispectral imagery and ground truth data. Later sections of this report review the investigation which addressed the feasibility of crop yield prediction using only the temporally-registered multispectral imagery.

---

*Note: Yield measurement is accomplished when the average crop yield of a site is measured by using registered multispectral images collected late in the growth cycle. The measurement is made after a stress map of the site is produced semiautomatically using the DICIFER software capability and ground truth data. Yield can be measured with apparent high accuracy by integrating the yield per pixel per stress class category over the entire digital stress map. Yield prediction is achieved by a different process using imagery collected earlier in the growth cycle.

The semiautomatic stress level discrimination effort utilized both image and ground truth data to assist in demonstrating the feasibility of semiautomatically mapping crop stresses as a function of tone measurements (image density) extracted from a specified month's three registered spectral images. Typically, the crop yield measurement study utilized each site's three registered spectral images of September or October and the results of the stress level mapping effort described previously. This effort was intended to demonstrate the feasibility of mapping yield as a function of mapped stress for which the yield of the crop per stress was known. The yield measurement investigation could indicate, for example, that by using October's multispectral images, a map of stress within the imaged site could be produced semiautomatically. Furthermore, if the mapped stress levels corresponded to stresses for which the yield had been measured at various minifield locations of a site, by integration of the yield per pixel assignments, a determination of the yield of the entire field could be obtained with some measure of accuracy.

Following this section's description of the image processing techniques utilized to semiautomatically map the levels of stress at Sites One, Three, and Seven, several examples of each site's stress maps will be presented. In conjunction with the

201

latter technical review, the procedures for yield measurement will be explained. The results of this study will be presented in the form of stress level maps, computer-generated yield measurements, and figures of accuracy.

Image Processing Procedures for Stress Level Discrimination

The theory behind the feasibility study for semiautomatically discriminating and mapping stresses within an agricultural crop is based upon laboratory and field experiments that have determined that a measure of a crop's health and potential yield is the amount of light that the crop's vegatative canopy reflects. In order to exploit this fact, the filters of the multispectral camera were selected to emphasize the stress levels of corn during the entire growth cycle. Consequently, the resultant digitized imagery contains density information for aiding a computer in semiautomatically mapping the stresses within an imaged corn field.

If the yield were sampled by measurement in various areas of a site which corresponded to different stressed areas of corn, a semiautomatically generated map of the stresses would also map the yield within the imaged site. That is, by using late season registered multispectral imagery of the corn field test site and a "design set", which consisted of computer area

202

files that bounded locations within the imaged site where yield
was sampled, a pattern recognition logic could be employed to
map out each of the sampled levels of stress of the site on a
pixel basis. After the decision image was created, which
effectively mapped the stresses within the imaged site, the
total number of pixels labelled a particular stress class
could be integrated and associated with the yield/pixel
assigned to the original sample area file of the design set.
After integration of each stress class's pixels and division
by the total number of pixels that occupied the field, a
measure in terms of bushels/acre could be made for the average
yield of the site.

Figure 66 outlines the basic image processing procedures
used for semiautomatically distinguishing and mapping levels of
stress. The discussion of the image processing effort should
begin with the block entitled, "Create Design Set." In creating
the design set, the interactive software capability of the
DICIFER System was utilized (pages 130-145). During this pro-
cess, graphic area files are drawn or specified by the user
around areas within a displayed image. Figure 61-63, page 184,
illustrated the locations of the design set area files of
Site One, Three, and Seven. Each of the black areas of the
three sites were the locations where yield had been measured
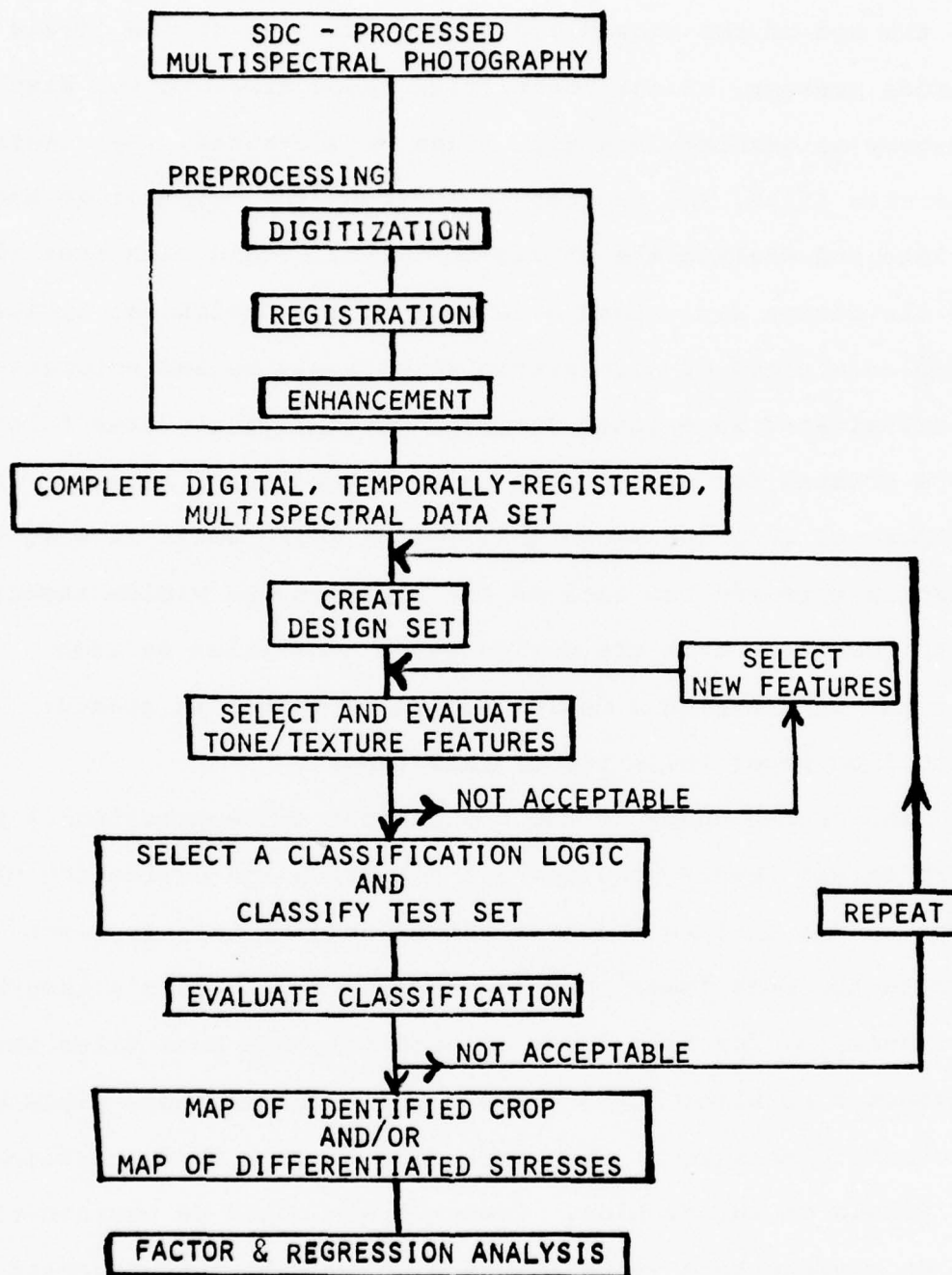
Figure 66. Outline of Digital Image Processing Procedures

at the end of the growth cycle as a function of the levels of stress present. After identifying these areas on the displayed imagery of October, box area files were created. By creating the area files, the user could instruct the computer to manipulate and analyze the pixels contained within each area file of the design set. Each area file of the design set typically sampled a class of corn stress which would be semiautomatically discriminated in a later stage of the analysis. Area files were created for each of the three sites' image sets for the purpose of sampling image information which would be used to discriminate and map each of the stresses and yields associated with locations that the design set's area files defined. This process of creating a design set is explained in greater detail on pages 180 - 188 of this report.

The second phase of the process for semiautomatically mapping stress levels requires the selection and evaluation of the tone and/or texture features which would be used to discriminate the area files' stress classes in each site's imagery. For example, for Site Three (Figure 62), six area files were presented as black boxes which occupied areas where yield was accurately measured. Each area file in Site Three contained 72 pixels of information. These pixels could be characterized by an average tone value (density, 0 to 255) and/or several

205

measures of texture. Furthermore, for one month's set of three multispectral images, the DICIFER capability permitted the extraction of an "infinite" number of features for each of the area files.

Tone features of any area file often included the mean density level of the 72 pixels of the green, red, and near-infrared images of any month's imagery (possible because the three images were registered). Moreover, the number of tone features for the entire image set totalled eighteen per stress class's area file.

Texture features could be specified by scanning each spectral image's area file with a small user-specified box which was usually dimensioned 3 x 3 pixels. Six different texture features could be computed (i.e., mean, standard-deviation, range, high density, low density, and median) for each nine pixel cluster scanned by the little box area within an area file's 72 pixels. The results of each series of computations could be averaged and the texture features could be assigned to the corresponding stress class. For example, if the latter process was performed on a single area file which defined pixels in only one spectral image of Site Three, six texture features could be extracted. Since texture varies as a function of the spectral image selected, a total of eighteen texture measures

could be developed for one month's data set of three images.
If all of the imagery of a given site were incorporated in the
design set of an area file of a stress class, the texture
features could number eighteen times the six texture features
per image or 108 texture features.

Infinite possibilities arise from the fact that features
can be added, subtracted, multiplied, divided, etc., to form
new measures for discriminating a class in a scene semiauto-
matically. The tone and texture features make up the infor-
mation which is extracted from images for discriminating the
different classes of stress or yield sampled by the area file.
The description of the DICIFER capability for feature extrac-
tion is presented on pages 133 to 136 of this report. Detailed
information of the feature extraction software is documented in
the reports of Forsen(18), Lietz(32) and Zanon(57).

Unfortunately, not all texture and tone features are good
for discriminating classes. In some images, two stresses may
have similar tone or texture characteristics. Consequently,
to use these particular features for stress discrimination
would confuse the computer and result in misclassification.
After the features are selected, they must be evaluated for
their ability to discriminate the classes. This step in the
image processing sequence is identified in the "Select and

Evaluate Features" step of Figure 66. By evaluating the
features selected for stress level discrimination using the
interactive DICIFER software, the Tektronix terminal, and the
Tektronix hardcopy capability, tone and/or texture features
can be either accepted or rejected from consideration. Those
which are rejected are not used in the semiautomatic classifi-
cation of stress. To include the latter feature would only
lead to confusion, poor stress maps, and yield measurement
inaccuracies. The excellent DICIFER software capability for
feature evaluation is described earlier in this report,
pages 136 - 138.

Figure 66 indicates that if the features are unsatisfactory
for stress level discrimination, an image processing loop is
available for recreating new measures or features. This might
be done by creating new area files and using new image data,
utilizing another month's data, or creating new features as
explained earlier. If the features are acceptable though, it
is possible to preliminarily evaluate the ability of the
features to discriminate classes by applying image classifi-
cation logic (i.e., Fisher logic, Boolean logic, One-or Two-
Space logic, etc.; Reference: pp. 138-140) to the pixels con-
tained in each of the area files of the design set. If any of
the features were poor stress discriminators, the resultant

"confusion matrix", Figure 67, would indicate this fact and steps could be taken to remedy the problem. The confusion matrix indicates the number of pixels contained within an area file of a particular stress class which were confused as another stress class. If there were no class confusions, a fair to good chance existed that the features would be good for classifying stress in the entire imaged scene.

If the design set's features prove to be satisfactory in the feature evaluation phase of the image processing procedures, the selected features become the criteria used to semiautomatically discriminate stress levels in the imaged corn field site or "test set". This step in the image processing effort is identified as the "Select a Classification Logic and Classify Test Set" step in Figure 66. The test set's pixels are classified by applying a classification logic which attempts to match each pixel's features to the diagnostic stress class features of the design set via the logic's decision criteria.

Though several varieties of logic are available for automated scene classification (i.e., Fisher, Boolean, One- or Two-Space, etc.), most of the semiautomatic stress level discrimination and mapping effort employed Fisher Pairwise logic. However, no matter which logic is utilized though, a semiautomatic classification is only as good as the features which are selected for the classification. Consequently, most of the

209

LOGIC FILE   VECTOR FILE

RATS   RATS

	A	B	C	D	E	F	R
A	42	0	0	0	32	0	0
B	0	57	15	0	0	0	0
C	0	30	42	0	0	0	0
D	0	0	0	72	0	0	0
E	45	0	2	0	25	0	0
F	33	0	0	0	39	0	0
R	2	0	0	0	0	0	0

LOGIC FILE   VECTOR FILE

DESN   DESN

	A	B	C	D	E	F	R
A	72	0	0	0	0	0	0
B	0	72	0	0	0	0	0
C	0	0	72	0	0	0	0
D	0	0	0	72	0	0	0
E	0	0	0	0	72	0	0
F	0	0	0	0	0	72	0
R	0	0	0	0	0	0	0

Figure 67. Two examples of confusion matrices utilized for
evaluation of the features selected for semiautomatic stress
level discrimination. The first example illustrates the class
confusion which occurs when the classifier logic is unable to
discriminate the stress classes using the selected design set
features. The second example illustrates the case of perfect
stress level discrimination.

work of this effort was concerned with varying the discriminating features for the purpose of improving the accuracy of distinguishing levels of stress. Comments concerning the semi-automatically-generated stress maps will emphasize the design set features, rather than the logic used to evaluate the test set. The logic will be simply specified. For a technical description of the Fisher, Boolean, and Nonlinear Mapping logic, reference should be made to this report's review of the DICIFER software, pages 136-140,or the literature (references: Bibliography of Zanon(57)).

The DICIFER software routines which are utilized interactively in the classification of a test set requests from the user three general categories of information: the design set features, the test set, and the logic to be implemented. In the specification of the design set, the area files which bound and sample imaged yield measurement locations are defined. After the area files are specified, the images to which the area files will be associated for feature extraction purposes are specified. The test set that will be semiautomatically classified is defined as either an area within each of the registered images or as the entire imaged scene. For example, suppose three stresses were to be classified and mapped based upon only the green, red, and near-infrared tone (density) features of the October registered-imagery for a given site. In specifying

211

the design set, the three area files corresponding to the yield-sampled stress locations would be input by name along with the images providing the features for the design set. The same three images, and any other area within the images that is considered the test set, would also be specified during the computer dialogue for test set specification. Finally, after the classifier logic is selected, the semiautomatic classification of the test set is performed. This final process consists of mathematically comparing the features defining each pixel of the test set with each of the features characterizing stresses defined by the design set. If the match is close enough, the pixel is classified as that particular stress class.

Classification of the test set pixels results in the creation of a decision image. The decision image provides a media for presenting the classified test set. Each density level of the decision image corresponds to a level of stress which was identified and mapped. For example, in the case of mapping three stresses, perhaps three density levels (i.e., 64, 128, or 255) would define the stress classes in the image and a "0" density would be used to represent pixels which were rejected during the classification; that is, pixels which could not be classified as one of the three levels of stress. A decision image can be displayed on television monitors or printed out on film hard copy using the DICIFER Color Printer.

212

This device color-codes each pixel on color film as a function of density level (i.e., 0-dark violet, 64-blue, 128-green, or 255-red). Figure 68 illustrates an example of a decision image which was output using the color printer. Each of the colors represents a particular level of stress which had image features in the design set that were determined to match those of the appropriately colored pixels of the decision image's test set.

Following the production of the decision image, the results are qualitatively evaluated by comparing the ground and image truth with the decision image's scene classification. If the decision image appears to be inaccurate, with the stresses and materials appearing to be misclassified, the decision image is rejected. Figure 66, page 204, indicated that the entire image processing cycle can be repeated in order to improve the accuracy of classification. However, if the decision image's stress map is acceptable, both the decision image and the count of the total number of pixels per stress class will be used for yield measurement. Examples of stress maps of Sites One, Three, and Seven will be presented along with pertinent information.*

---

*Note: Since the overall objective of this study was to investigate the feasibility of crop yield prediction, elaborate detail and explanation of each decision image is avoided because this would tend to distract the reader from the objectives of the effort.

Figure 68. Illustration of a semiautomatically-generated
decision image of Site One. Stresses were mapped semiauto-
matically using features extracted from registered, multi-
spectral images. The chart below indicates the number of
pixels identified as a particular level of stress within the
imaged field. The pixel count per stress class was obtained
using a DICIFER software routine.

Decision Image Pixel Count per Stress Class
Image Name-POPS, Site One

Stress Class	Number of Pixels	Decision Image Coded-Density	Color Code
A	129	164	Yellow
B	1149	100	Blue-Green
C	1421	255	Red
D	301	32	Light Purple
R (Rejects)	1489	0	Dark Purple
Sum:	4489		

214

Yield Measurement

After a site's stress map was generated, the yield of the entire field was measurable by using the pixel count per each stress class of the decision image. Since each area file had been selected where both the stress and yield were known, the semiautomatically-generated stress maps (decision images) could be considered maps of yield within the site.* Figure 69 presents an example of how yield was actually measured from a decision image's stress map of Site One, <u>POPS</u>, Figure 68.

The technicalities of measuring yield using decision images requires that the pixel count per stress class be known and the yield measured for each stress class be known. This criteria was satisfied and yield measurements were obtained with the following steps. Original measurements of yield were collected in four sites of Site One (reference the Ground Truth description, pages 65-73). Table 5, page 83, indicates the

---

*Note: This apparent axiom or assumption of the capability of the image data set is supported by studies which have linked stress with yield, stress with a crop canopy's albedo, and yield potential of a crop with the crop canopy's albedo (reference the literature review of crop yield prediction investigations, pages 25-39). Furthermore, though it will be demonstrated that accurate measurement of a site's yield can be obtained using the decision images, there is no quantitative evidence that the stress maps were accurate maps of yield patterns within a field. The validity of both the axiom and the accuracy of the stress map will need to be established through continued experimentation on a larger scale.

215

Figure 69. Location of the yield measurement areas of Site One. The numbers correspond to the yield measurement locations (Table 5 , page 83) and the letters correspond to area file names and stress class codes (Table 13 , page 219). Using POPS decision image, Figure 68, yield measurement was performed in the following manner:

Stress Class	Pixels Per Class	Multiply	Ground Truth Yield (Bu/Acre)	Product if "R Class Yield is: 0 bu/a	124 bu/a
R	1489	X	0/124	0	184636
D	301	X	157.0	47257	47257
C	1421	X	123.9	176062	176062
B	1149	X	107.6	123632	123632
A	129	X	129.4	16693	16693
Sum:	4489		Summation:	363644	548280

Division by Total Pixels of Site One: 4489 4489

Results: Average Yield of Site One: 81 bu/a 122.0 bu/a

Independent Measure of Yield, Site One: 124 bu/a 124 bu/a

Accuracy of Computer Measurement: 65% 98%

Note: If total yield of the field is required (bushels), accurate mensuration is required to determine the true acreage of the field. However, since average yield is the requirement, there is no need to perform accurate mensuration. There is a requirement though, to insure that each area file, defining a stressed area where yield had been measured, has been accurately located in the design set. Further note that this computation is only an example and should not be construed to be the ideal example of yield measurement for Site One.

216

yield for each sample location of Site One. DICIFER Software provided a routine for counting the number of pixels per stress identified in the decision image. Since a pixel images an area of ground which can be associated with a measurable yield, a pixel can be translated into a yield measurement (i.e., bushels per acre). If the yield associated with each pixel of a decision image is integrated for all the pixels in the imaged field, the resultant quantity is a measure of the field's total harvest. Finally, by dividing the total yield of the field by the total number of pixels which define the field, a measure of the site's average yield can be obtained. This sequence of steps is illustrated in Figure 69.

Yield measurement was performed on most of the decision images generated for the three sites. For obtaining figures of yield measurement accuracy, the computed value was compared to the average yield of the field quoted by an independent information source - the site manager. The site manager measured the yield using the conventional farming practices for yield measurement.*

_____

*Note: Three sources of yield measurement were provided during the course of the study: 1) ground truth survey prior to harvest; 2) the site manager's survey after harvest; 3) yield measurement provided by the County Extension Service's 5-Acre Corn Sample Yield Estimate. Generally, the site manager's estimate was most accurate (though in some instances, it was obvious to the investigators that the site manager preferred to state the yield he would have had rather than what was actually harvested).

217

The following examples, Figures 70-96, illustrate the different techniques investigated for semiautomatically classifying Sites One, Three, and Seven's stresses. For those images which seemed to be reasonable stress maps of the site, the yield was calculated using the techniques presented in Figure 69. The computed yield measurement and accuracy is listed with each example's decision image. For those examples with particularly interesting results, a short paragraph commenting about the technicalities of the classification was included in the caption for the figure. The reader **should** reference the examples of the additive color images of the three sites, Figures 32,37&40,to qualitatively evaluate the logic and **features' capability to discriminate levels of stress.**

Tables 13-15 precede the decision images of each site so that some understanding can be established of which colors of the decision images correspond to specific stresses and yields. Due to difficulties in reproducing colors in a report using third generation negatives, the colors will not be uniform from image to image. Generally, a color can be correlated to a specific yield which was measured in a specific location of the field and mapped throughout the imaged scene. Conclusions of this effort are presented following the presentation of the three sites' respective decision images and yield measurements.

Site One

Ground Truth Yield Measurement Location & Stress Class Symbol	Decision Image Density	Expected Color	Yield (Bu/Acre)
1 / C	255	Red	129.4
2 / A	164	Yellow	107.6
3 / B	100	Lt Blue-Lt Green	123.9
4 / D	32	Lt Purple-Lt Blue	157.0
Rejects: R	0	Dark Purple	0 or 124

Table 13. This table lists the expected yields associated with the colors of the mapped stresses within the decision images of Site One. The colors may not necessarily hold true due to limitations in color reproduction during the report publication. Consequently, the expected color per stress is expected as a range of color (i.e., Light Blue to Light Green). Reference Figure 69 for identifying the yield measurement locations of Site One.

Figure 70.   Image Name-<u>NEXT</u>; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 24
Spectral images utilized: B,G,R,IR (May thru October)
Number and dimensions of area files: 4, (4) pixels/area file
Features: Tone (image density or grey level of pixels)
Classification Logic: Boolean

Note: This decision image, which was produced early in this
study, utilized **area** files which enclosed the imaged loca-
tions of the site in which the yield was actually measured.
However, the Foley criteria dictated the use of area files
which contained 72 pixels. Poor classification results forced
the area files to be redimensioned to match the specifications
of the Foley criteria. The Boolean classifier logic used tone
features which were +3 standard deviations of the mean density
of the YSAX value of the area file.

Figure 71.   Image Name-<u>NEWT</u>; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 24
Spectral images utilized: B,G,R,IR (May thru October)
Number and dimensions of area files: 4, (72) pixels/area file
Features: Tone
Classification Logic: Boolean

Note: Foley criteria utilized in dimensioning area files.
Boolean classifier used tone features within the range of
+2 standard deviations of the mean density value of the area
file pixels for each spectral image of the design set.

220

Figure 72.   Image Name-LUST; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 24
Spectral images utilized: B,G,R,IR (May thru October)
Number and dimensions of area files: 4, (72) pixels/area file
Features: Tone
Classification Logic: Fisher

1)   Computed yield measurement: 123.5 bu/acre
2)   Ground truth average yield measurement: 129.4 bu/acre
3)   Independent average yield measurement: 124 bu/acre
        Accuracy: 1 vs 3 = 99%

Figure 73.   Image Name-XXXX; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 24
Spectral images utilized: B,G,R,IR (May thru October)
Number and dimensions of area files: 4, (72) pixels/area file
Features: Tone Residual*
Classification Logic: Fisher

1)   Computer yield measurement: 121.5 bu acre
2)   Ground truth average yield measurement: 129.4 bu/acre
3)   Independent average yield measurement: 124 bu/acre
        Accuracy: 1 vs 3 = 98%

_____

*Note: Tone residual calculated from formula:
$M = ((D - GLTX) + 255)/2$, where (M) is the new residual
tone value calculated for each pixel having an original
density (D) in the design and test sets. (GLTX) is the
mean density value of the pixels of the imaged site in
each spectral image of the image data set (Ref: Conclusions.
page 237-40).

Figure 74.   Image Name-KEAN; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 3
Spectral images utilized: G,R,IR (October)
Number and dimensions of area files: 4, (72) pixels/area file
Features: Tone
Classification Logic: Fisher

1)   Computed yield measurement: 126.8 to 129.8 bu/acre
2)   Ground truth yield measurement: 129.4 bu/acre
3)   Independent average **yield measurement: 124 bu/acre**
     Accuracy: 1 vs 3 = 96%

Note: Compare this decision image to that of Figure 72 which
maps the levels of stress based upon the pixel's spectral
reflectance history over the entire growth cycle (24 images,
instead of 3).   Reference comments in Conclusions, page 240-42.

Figure 75.   Image Name-SEAN; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 3
Spectral images utilized: G,R,IR (October)
Number and dimensions of area files: 4, (72) pixels/area file
Features: Tone Residual (Reference: Caption of Figure 73)
Classification Logic: Fisher

1)   Computed yield measurement: 128.3 bu/acre
2)   Ground truth yield measurement: 129.4 bu/acre
3)   Independent average yield measurement: 124 bu/acre
     Accuracy: 1 vs 3 = 97%

222

Figure 76.   Image Name-POKE; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 24
Spectral images utilized: B,G,R,IR (May thru October)
Number and dimensions of area files: 4, (72) pixels/area file
Features: Tone
Classification Logic: Nonlinear Mapping

Note: Nonlinear mapping logic, which is a supervised
clustering logic typically used for analyzing the structure
of the data, was utilized to identify natural stress patterns
within the imaged Site One. Unfortunately, this logic can
only discriminate 3 classes at a time. The decision image is
the result of the 33rd iteration of clustering. This iteration
was selected for display because the class discrimination
improved until this iteration; following this iteration,
classes began to converge.

Figure 77.   Image Name-POPS; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 24
Spectral images utilized: B,G,R,IR (May thru October)
Number and dimensions of area files: 4, (72) pixels/area file
Features: Tone-Variance; Calculated with the formula:
        $M = (D - GLTX)***3$, where the new cubed residual
value (M) is **calculated** for each pixel tone value (D) in the
design and test sets. (GLTX) is the mean density value of the
pixels of the imaged site in each spectral image of the image
data set.
Classification Logic: Fisher

Note: This decision image resulted from an attempt to measure
yield with a special tone feature derived from this study's
theory, page **189** . Unfortunately, the PDP-11/20 minicomputer
truncates any calcualtion greater than 255 to 255 and any value
less than 0 to 0. Reference Figure 69 for yield measurement
computations.

223

Figure 78.   Image Name-MEAN; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 1
Spectral images utilized: Near-infrared (October)
Number and dimensions of area files: 4, (72) pixels/area file
Features: Tone
Classification Logic: Fisher

1)   Computed yield measurement: 126.3 bu/acre
2)   Ground truth yield measurement: 129.4 bu/acre
3)   Independent average yield measurement: 124 bu/acre
       Accuracy: 1 vs 3 = 98%
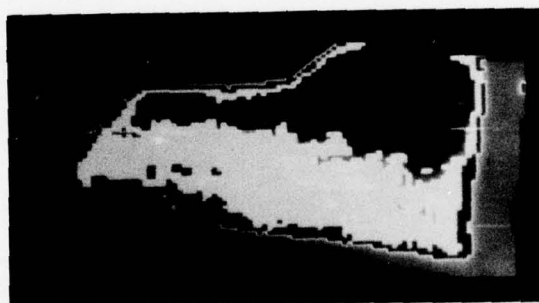
Figure 79.   Image Name-TEAM; Digital Stress Map of Site One

Number of spectral images used in the design and test sets: 1
Spectral images utilized: Near-infrared (October)
Number and dimensions of area files: 4, (72) pixels/area file
Features: Tone Residual (Reference: Caption Figure 73)
Classification Logic: Fisher

1)   Computed yield measurement: 126.4 bu/acre
2)   Ground truth yield measurement: 129.4 bu/acre
3)   Independent average yield measurement: 124 bu/acre
       Accuracy: 1 vs 3 = 98%

Note: Neither Figures 78 or 79 seem acceptable for yield
measurement, though accuracies appear to be high.   The
decision images developed from three images of October provide
the most reasonable stress maps of Site One, **Figures 74 & 75.**

224

Figure 80. Location of the yield measurement areas of Site
Three. The numbers correspond to the yield measurement
locations (Table 6 , page 96) and the letters correspond
to area file names and stress class codes (Table 14 , next page).
Pixel information in each area file was utilized as features
for semiautomatically mapping yield and stress in Site Three's
imagery.

Note: Referencing Figure 80 and Figures 57&58 (page 177),
two area files were sampled in locations where a portion of
Site Three's temporally-registered multispectral imagery had to
be deleted because of cloud cover problems in July and dif-
ficulties in combining September's imagery (Ref: Figure 80
w/respect to Figures 57 & 58). Consequently, some of the
decision images illustrated were mapped using the four area
files which were not affected by the image processing problem.
Other decision images were produced using six area files, two
of which were selected in areas of the image that had similar
spectral characteristics of the two original, but unuseable
area files (i.e., Figures 81 & 83).

225

Site Three

Ground Truth Yield Measurement Location & Stress Class Symbol	Decision Image Density	Expected Color	Yield (Bu/Acre)
1 / A	255	Red	91.5
2 / B or Y	220	Orange	78.4
3 / C	180	Yellow	0.0
4 / D or X	120	Light to Dark Green	120.0
5 / E	80	Light Blue to Light Green	95.1
6 / F	30	Light Purple to Light Blue	62.4
Rejects: R	0	Dark Purple	0.0 / 90.0

Table 14: This table lists the expected yields associated with the colors of the mapped stresses within the decision images of Site Three. The colors may not necessarily hold true due to limitations in reproducing the color during the report publication. Consequently, the expected color per stress is expressed as a range of color (i.e., Light blue to Light green). Reference Figure 80 for identifying the yield measurement locations of Site Three.

Figure 81. Image Name-YUMM; Digital Stress Map of Site Three

Number of spectral images used in the design and test sets: 18
Spectral images utilized: G,R,IR (May thru October)
Number and dimensions of area files: 6*, (72) pixels/area file
Features: Tone
Classification Logic: Fisher (4 and 5 vote criteria gave different classifications, both poor)

1) Computed yield measurement: 67 to 79 bu/acre
2) Ground truth yield measurement: 89.7 bu/acre
3) Independent average yield measurement: 90 bu/acre
     Accuracy: 1 vs 3 = 80+6%, range varies with interpretation
of reject class yield (0 or 90 bu/a)

Note: Because of cloud cover and image combination problems,
two of the original areas were not useable. By interpreting
density maps of the site (line printer output) and ground
truth data, two areas were selected which approximated the
features of the two original areas. The resultant stress map
above is different from those produced from October's three
images. The former map probably represents pixels within the
field which had a similar reflectance history and yield; whereas,
the October map may be a short term stress level map. However,
due to the ambiguity of the two new area files, Figure 81 is
probably not a reliable yield map, despite the relatively good
accuracy figures.

Figure 82. Image Name-CUMM; Digital Stress Map of Site Three

Number of Spectral images used in the design and test sets: 18
Spectral images utilized: G,R,IR (May thru October)
Number and dimensions of area files: 6,(Ref: Fig. 81 comments)
Features: Tone Residual (Reference: Caption, Figure 73)
Classification Logic: Fisher (4 and 5 vote criteria gave
different classifications, both poor)

1) Computed yield measurement: 79 to 84 bu/acre
2) Ground truth yield measurement: 89.7 bu/acre
3) Independent average yield measurement: 90 bu/acre
     Accuracy: 1 vs 3 = 90 ± 3%, range varies with interpre-
tation of reject class yield (0 or 90 bu/a).

227
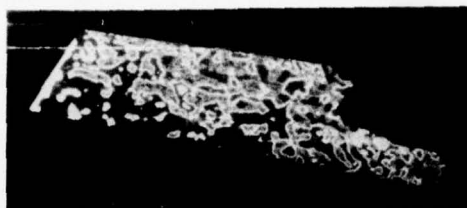
Figure 83.   Image Name-PIGY; Digital Stress Map of Site Three

Number of spectral images used in the design and test sets: 1
Spectral images utilized: Near-infrared (September)
Number and dimesnions of area files: 6, (72) pixels/area file
(Ref: Area file comments, Fig: 81)
Features: Tone
Classification Logic: Fisher (4 & 5 vote criteria gave similar
classification results)

1)   Computed yield measurement: 93 bu/acre
2)   Ground truth yield measurement: 89.7 bu/acre
3)   Independent average yield measurement: 90 bu/acre
     Accuracy: 1 vs 3 = 96%

Note: This decision image's stress pattern corresponds with
those of the additive color imagery of Site Three, Figure 37 ,
page 84 . Referencing the image truth photography, the pattern
seems to correspond with different exposed horizons of soil
which certainly have an effect on yield and photosynthesis.
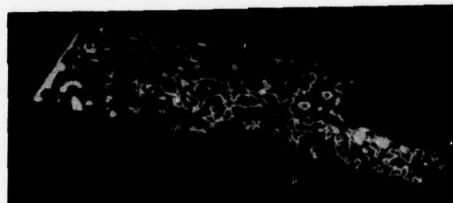
Figure 84.   Image Name-MORN; Digital Stress Map of Site Three

Number of spectral images used in the design and test sets: 1
Spectral images utilized: Near-infrared (September)
Number and dimensions of area files: 6, (72)pixels/area file
(Ref: Area file comments, Fig: 81)
Features: **Tone** Residual (Reference: Caption, Figure 73)
Classification Logic: Fisher (4 & 5 vote criteria gave similar
classification results)

1)   Computed yield measurement: 92 bu/acre
2)   Ground truth yield measurement: 89.7 bu/acre
3)   Independent average yield measurement: 90 bu/acre
     Accuracy: 1 vs 3 = 97.7%

Note: Comparing Figures 83 & 84 it appears that the tone
features and tone residual features give similar stress classifi-
cation and yield measurement results. This would not occur if
digitization was grossly nonlinear as a result of the Computer
Eye.

228
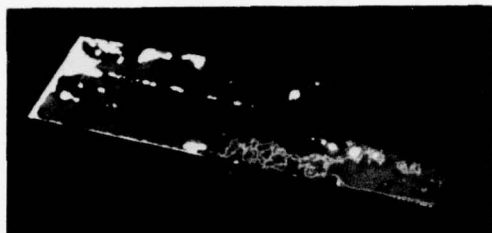
Figure 85.   Image Name-RICK; Digital Stress Map of Site Three

Number of spectral images used in the design and test sets: 1
Spectral images utilized: Near-infrared (October)
Number and dimensions of area files: 6, (72) pixels/area file
(Original 6 area files used)
Features: Tone
Classification Logic: Fisher (4 & 5 vote criteria gave smiliar
stress classification results)

1)   Computed yield measurement: 85 bu/acre
2)   Ground truth yield measurement: 89.7 bu/acre
3)   Independent average yield measurement: 90 bu/acre
     Accuracy: 1 vs 3 = 94.4%

Figure 86.   Image Name-MICK; Digital Stress Map of Site Three

Number of spectral images used in the design and test sets: 1
Spectral images utilized: Near-infrared (October)
Number and dimensions of area files: 6, (72) pixels/area file
(Original 6 area files used)
Features: Tone Residual (Reference: Caption, Figure 73)
Classification Logic: Fisher (4 & 5 vote criteria gave
similar stress classification results)

1)   Computed yield measurement: 87 bu/acre
2)   Ground truth yield measurement: 89.7 bu/acre
3)   Independent average yield measurement: 90 bu/acre
     Accuracy: 1 vs 3 = 96.6%

Note: Figures 85 and 86 appear to be the same but the pixel
counts per stress indicate that there were some differences in
stress classification.

Figure 87.   Image Name-ROBN; Digital Stress Map of Site Three

Number of spectral images used in the design and test sets: 3
Spectral images utilized: G,R,IR (October)
Number and dimensions of area files: 6, (72) pixels/area file
(Original 6 area files used)
Features: Tone Residual (Reference: Caption, Figure 73)
Classification Logic: Fisher (4 & 5 vote criteria gave similar
stress classification results)

1)   Computed yield measurement: 90.5 to 92.7 bu/acre
2)   Ground truth yield measurement: 89.7 bu/acre
3)   Independent average yield measurement: 90 bu/acre
     Accuracy: 1 vs 3 = 98%

Note: Though this decision image did illustrate discrimination
of weeds, it primarily maps levels of stress within the field.
This means that a particular color does not necessarily repre-
sent a stress which can be identified with a particular disease.
The colors represent a level of efficiency at which the crop is
carrying on photosynthesis!
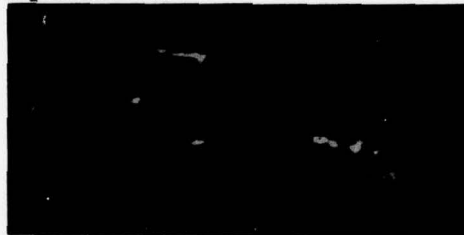
Figure 88.   Image Name-DEBB; Digital Stress Map of Site Three

Number of spectral images used in the design and test sites: 3
Spectral images utilized: G,R,IR (October)
Number and dimensions of area files: 6, (72) pixels/area file
(Original 6 area files used)
Features: Tone
Classification Logic: Fisher (4 & 5 vote criteria gave similar
stress classification results)

1)   Computed yield measurement: 91 to 93 bu/acre
2)   Ground truth yield measurement: 89.7 bu/acre
3)   Independent average yield measurement: 90 bu/acre
     Accuracy: 1 vs 3 = 98+%

Note: Figures 87 & 88 appear to be the same but the pixel
counts per stress indicate that there were differences in
classification. In most cases, the tone features and tone
residual features obtained the same decision image if the same
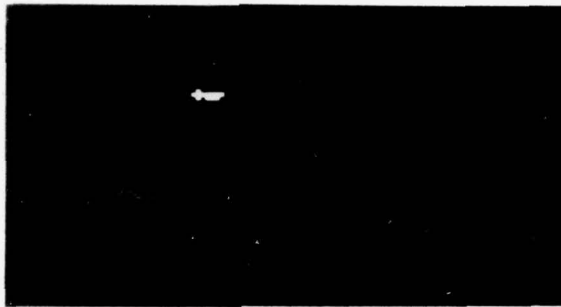logic was utilized.

230
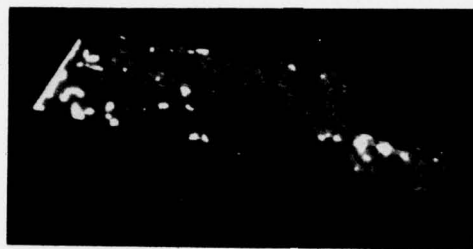
Figure 89. Image Name-WWWW; Digital Stress Map of Site Three

Number of spectral images used in the design and test sets: 1
Spectral images utilized: Near-infrared (September)
Number and dimensions of area files: 4, (72) pixels/area file
(Original 4 of 6 area files used)
Features: Tone Residual (Reference: Caption, Figure 73)
Classification Logic: Fisher

1) Computed yield measurement: 80 bu/acre
2) Ground truth yield measurement: 89.7 bu/acre
3) Independent average yield measurement: 90 bu/acre
      Accuracy: 1 vs 3 = 88%

Note: When these same stress mapping procedures were followed
using tone features, the resultant decision image was
identical to Figure 89.

231
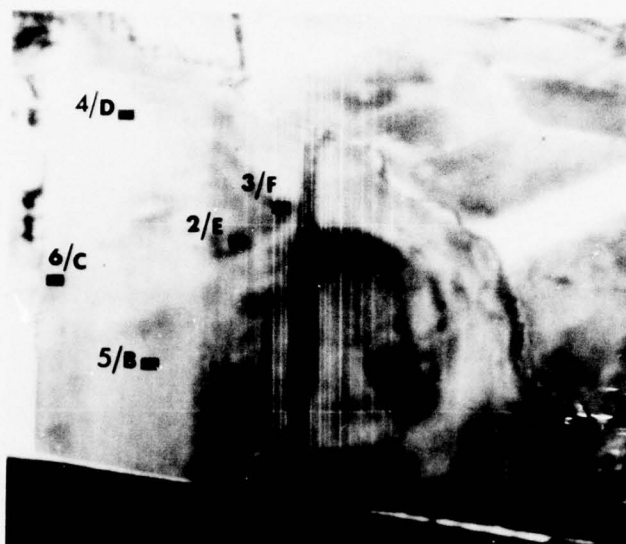
Figure 90. Location of the yield measurement areas of Site
Seven. The numbers correspond to the yield measurement loca-
tions (Table 7 , page111) and the letters correspond to area
file names and stress class codes (Table 15, next page). Pixel
information in each area file was utilized as features for
semiautomatically mapping yield and stress in Site Seven's
imagery.

Note: Reference Figure 60, page 179, two area files' locations
were sampled in locations where a portion of Site Seven's tem-
porally-registered, multispectral imagery had to be rejected
from consideration due to problems associated within combining
the imagery of October. Consequently, depending upon the
month's or months' imagery being used to map stresses, the area
files used for feature extraction may number from 4 to 6. The
final stress maps were generated with September's imagery which
utilized all six area files data.

Site Seven

Ground Truth Yield Measurement Location & Stress Class Symbol	Decision Image Density	Expected Color	Yield (Bu/Acre)
1 / A	225	Red	124.2
2 / E	64	Dk to Med Blue	115.8
3 / F	32	Lt Purple to Lt Blue	108.6
4 / D	100	Lt Blue to Lt Green	142.3
5 / B	210	Orange to Brown	130.2
6 / C	160	Yellow	87.9
Rejects: R	0	Dk Purple	0 / 125

Table 15. This table lists the expected yields associated with the colors of the mapped stresses within the decision images of Site Seven. The colors may not necessarily hold true due to limitations in reproducing the color during the report publication. Consequently, the expected color per stress is expressed as a range of color (i.e., Light blue to Light green). Reference Figure 90 for identifying the yield measurement locations of Site Seven.

Figure 91.   Image Name-DANC; Digital Stress Map of Site Seven

Number of spectral images used in the design and test sets: 3
Spectral images utilized: G,R,IR (September)
Number and dimensions of area files: 6, (72) pixels/area file
Features: Tone
Classification Logic: Fisher (4 & 5 vote criteria gave similar stress classification results)

1)   Computed yield measurement: 119 to 124.8 bu/acre
2)   Ground truth yield measurement: 118 bu/acre
3)   Independent average yield measurement: 125 bu/acre
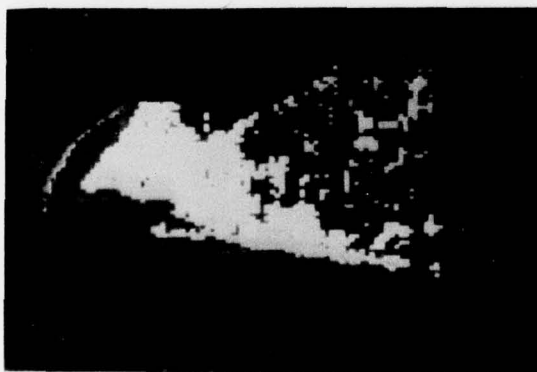   Accuracy: 1 vs 3 = 97.5+2%

Figure 92.   Image Name-REST; Digital Stress Map of Site Seven

Number of spectral images used in the design and test sets: 3
Spectral images utilized: G,R,IR (September)
Number and dimensions of area files: 6, (72) pixels/area file
Features: Tone Residual (Reference: Caption, Figure 73)
Classification Logic: Fisher (4 & 5 vote criteria gave similar stress classification results)

1)   Computed yield measurement: 119 to 126 bu/acre
2)   Ground truth yield measurement: 118 bu/acre
3)   Independent average yield measurement: 125 bu/acre
       Accuracy: 1 vs 3 = 97+2%

234

Figure 93. Image Name-IMAG; Digital Stress Map of Site Seven

Number of spectral images used in the design and test sets: 1
Spectral images utilized: Near-infrared (September)
Number and dimensions of area files: 6, (72) pixels/area file
Features: Tone
Classification Logic: Fisher (4 & 5 vote criteria gave similar
stress classification results)

1) Computed yield measurement: 122.3 bu/acre
2) Ground truth yield measurement: 118 bu/acre
3) Independent average yield measurement: 125 bu/acre
     Accuracy: 1 vs 3 = 98%
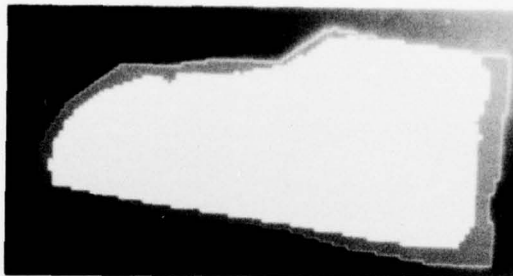

Figure 94. Image Name-FISH; Digital Stress Map of Site Seven

Number of spectral images used in the design and test sets: 1
Spectral images utilized: Near-infrared (September)
Number and dimensions of area files: 6, (72) pixels/area file
Features: Tone Residual (Reference: Caption, Figure 73)
Classification Logic: Fisher (4 vote criteria)

1) Computed yield measurement: 121.6 bu/acre
2) Ground truth yield measurement: 118 bu/acre
3) Independent average yield measurement: 125 bu/acre
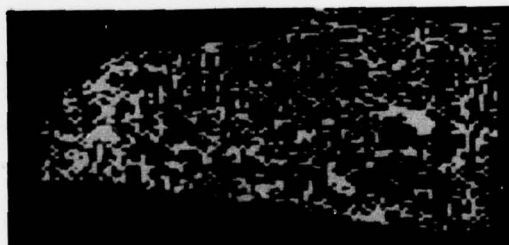   Accuracy: 1 vs 3 = 97%

235
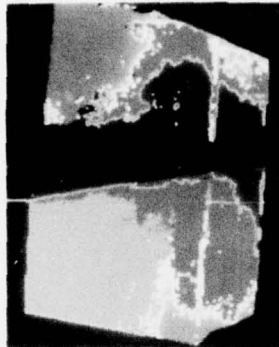
Figure 95.   Image Name-COOK; Digital Stress Map of Site Seven

Number of spectral images used in the design and test sets: 3
Spectral images utilized: G,R,IR (October)
Number and dimensions of area files: 4, (72) pixels/area file
(Reference: Caption, Figure 90)
Features: Tone
Classification Logic: Fisher (5 vote criteria)

1) Computed yield measurement: 125 to 127 bu/acre
2) Ground truth yield measurement: 118 bu/acre
3) Independent average yield measurement: 125 bu/acre
    Accuracy: 1 vs 3 = 99+1%

Note: There are general similarities between the stress
classification of October's imagery and September's imagery
(Figure 91). Differences occur because a different number of
area files were used in the design set of this classification.
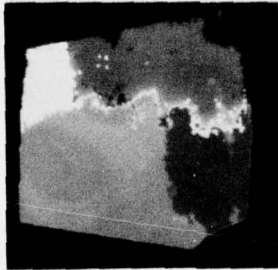
Figure 96.   Image Name-PLUK; Digital Stress Map of Site Seven

Number of spectral images used in the design and test sets: 15
Spectral images utilized: G,R,IR (May thru September)
Number and dimensions of area files: 5, (72) pixels/area file
(Reference: Caption, Figure 90)
Features: Tone
Classification Logic: Fisher (4 vote criteria)

1) Computed yield measurement: 118.7 to 121.0 bu/acre
2) Ground truth yield measurement: 118 bu/acre
3) Independent average yield measurement: 125 bu/acre
   Accuracy: 1 vs 3 = 96+1% (Reference: Caption, Figure 81)

236

Conclusions

Stress maps were semiautomatically generated using different types of image classifier logic, feature design sets, and test sets. Though the principal test set consisted of October's spectral images, other sets of images were also studied for their potential as data sets for stress mapping and yield measurement. These other test sets included both September's spectral images and sets of months' imagery (i.e., April thru October, April thru September). The conclusions of this effort report the observed performances of the classification logic, the feature design sets, and the test sets utilized for stress level mapping and yield measurement. The following conclusions are based on many stress mapping and yield measurement experiments; the illustrations preceding this section represents only a cross-section of those experiments.

Logic

Boolean logic, Nonlinear Mapping logic (supervised clustering), and Fisher Discriminant logic were utilized during this study of semiautomatic stress level mapping. The results of the Boolean and Nonlinear Mapping experiments with Site One's imagery, Figures 70,71 & 76, illustrate the reason why most of the experiments were performed with the Fisher logic. Though Fisher logic was used a great deal, it is certainly not

237

identified as an optimal criteria for stress mapping. Fisher

logic does not permit the bounding of classes exactly enough

in the n-dimensional feature space. Fortunately, there are

other types of logic that attempt to bound classes more exactly

by containing a class's features in hypervolumes instead of

hyperplanes in the n-dimensional feature space (i.e.,

Mahalanobian versus Fisher). Nevertheless, Fisher logic was

fast, easy to implement, and generated reasonably good stress

maps with each site's imagery.

Features Utilized for Classification

Though many different features were investigated for semi-

automatic stress mapping purposes, there were basically three

categories of features that were principally utilized in the

creation of the design and test sets: tone features, tone-

residual features, and tone-variance features. The selection of

these feature categories was governed by the criteria of the

yield prediction feasibility study's theory which required

tone-related, image-extracted data. The image-extracted infor-

mation that would be used in the **yield** prediction model would

be either the average density of each area file's pixels or a

tone-residual or variance measure calculated between each area

file's pixels and the mean density level of the imaged site

(Reference: Figures 64 & 65, pages **187**& **195**, respectively).

238

Furthermore, the latter tone-residual and tone-variance features were created for the purposes of nullifying the problems of nonlinear digitization, unequal density ranges, etc., which are described in the crop yield prediction theory section of this report (pages 189-199).

The first specialized tone feature created was the tone-variance measure (not a true density variance measure, it might be better named a "tone-residual-cubed" feature). This feature was calculated for each pixel of the design and test set with the formula:

$M = (D - GLTX)^3$, where "D" is the density level of each pixel of the design and test sets; and "GLTX" is the mean density level of the pixels contained within the boundaries of the imaged site for each spectral image of the test set; and "M" is the new cubed residual or density "variance" value of the pixel. The "M" value was to be the feature utilized for classifying stress and for predicting yield.

Figure 68 illustrated the results of using the latter features for classifying stress in Site One's imagery. The disappointing problem was that the PDP-11/20 minicomputer was truncating all calculation results which were greater or less than "255" or "0" to either "255" or "0" respectively. This naturally resulted in many stress misclassifications. Consequently, a new specialized tone feature was developed which simply computed the tone-residual value between each pixel and the mean density of the imaged site with the formula:

239

M = ((D - GLTX) + 255)/2, where "D" is the density
level of each pixel of the design and test sets; "GLTX"
is the mean density level of the pixels contained within
the boundaries of the imaged site for each spectral
image of the test set; and "M" is the new density level
of the subject pixels. This "M" value was used as the
feature for classifying stress levels and for inves-
tigating the feasibility of predicting yield. Note
that the formula produces new values which are always
within the 0 to 255 range.

Both tone and tone-residual features demonstrated excellent semiautomatic stress level classification results. The two types of feature categories were compared by using the same month(s)' imagery and design sets. The resultant decision images were similar stress maps which generated comparable yield measurement accuracies (i.e., Figures 72 & 73; 74 & 75; 81 & 82; 85 & 86; 92 & 93, etc). For the purpose of this feasibility study, these features were judged to be adequate for stress level mapping. However, it should not be assumed that any of the features used for stress mapping were optimal. Optimal features could be established during another study of larger scale.

Yield Measurement With Respect to Test Sets

Yield measurement was performed with each site's October imagery. However, yield measurement was also accomplished with some site's September imagery, April thru October imagery, and April thru September imagery. After analyzing the yield

240

measurements derived from the three sites' stress maps, the following conclusions can be made:

1) The most accurate stress level maps of each site, which provided the most consistent high accuracies for yield measurement (i.e., 96 to 98%), were produced when either September or October's three spectral images (i.e., green, red, and near-infrared) were utilized as a data source for tone features for the design and test sets.

2) Reasonable maps of stress levels, demonstrating relatively high yield measurement accuracies, were obtained when only the September or October's near-infrared image of a site was utilized for features of the design and test sets. However, comparing the decision images or stress maps to the ground truth, it appears that the most accurate maps of stress levels were demonstrated when either September or October's three spectral images were used.

3) For those decision images which were generated from the three spectral images of the months of April thru October or September, Figures 72,73,81,82 & 91, the ability to measure yield was inconsistent (85% to 98% accuracy). The reason for the range of accuracies is believed to have been caused by the ambiguity of the substitute area files used in some of the design sets of Site Three (Figure 81). The investigator interprets these decision maps as representing the stressed

241

corn having similar reflectance histories and yields (comment
is supported by the literature concerning crop yield prediction,
pages 25-39). These decision images are quite distinct from
those generated from only one month's imagery. Perhaps the
stress maps of September or October represent short term stress
maps. If the latter statement is true, the decision images
which are based on several months' imagery may provide the
optimal yield measurement data set.

Multivariate Analysis

Prior to investigating the feasibility of crop yield pre-
diction using regression analysis techniques, a multivariate
(factor) analysis was performed for determining the underlying
structure of the ground truth and image-extracted data.
Approximately eighty different categories of data were col-
lected during the course of this feasibility study in crop
yield prediction. Determination of the data's structure would
identify which variables account for the greatest proportion
of the total variability of the data (variance). Since the
significant variables would be incorporated into a linear
regression analysis for demonstrating the feasibility of
yield prediction, it was important to determine which remotely
detectable variables were both independent and correlated well
with the yield variables.

Research Question

Multivariate analysis might be described as a mathematical
technique that seeks to uncover the underlying dimensionality
of the data(Yeates,56). Many social scientists have utilized
the techniques of multivariate analysis in identifying both
significant variables and relationships. Specifically,
multivariate analysis determines which variables make up the
basic information structure of a relatively large number of

243

variables having measures which overlap. For example, factor analysis, which is a type of multivariate analysis, computes hypothetical variables or "factors" which have the attribute of accounting for the variance originally defined by the inter-correlations of a larger set of measures(Williams,55).

An objective was formulated to guide the efforts of the multivariate analysis of the eighty variables' data collected during the ground truth and imagery acquisition programs of this feasibility study. The variables could be divided into three categories: image-extracted variables, weather variables, and ground truth variables. Since the ultimate goal of this study was to determine the feasibility of predicting the yield of a crop by using remotely-sensed data, it was important to determine which of the variables or class of variables best accounted for the total variability of the data set acquired from the three agricultural test sites. It was concluded that the research question should concern itself with the deter-mination of which variables, either remotely-sensed or those derived from ground truth, accounted for most of the variance of all of the data. The analysis would identify specific variables with "factors" (independent hypervariables defining the actual structure of the data). Such an output would be useful in determining which variables should be included in the regression analyses. This research utilized the principal

244

component and factor analysis techniques to also aid in the determination of the significance of the spectrally and temporally variant nature of the image data relative to the weather and ground truth data.

Factor and Principal Component Multivariate Analysis

Two general types of multivariate analysis were utilized during the course of this study; principal component analysis and factor analysis. Though both principal component analysis and factor analysis studies were conducted at the beginning of this effort, due to the similarity in results between the two analysis techniques*, the principal component analysis was emphasized in the interpretation section.

Possibly the best way of expressing multivariate analysis theory is to begin the discussion with a geometric conceptualization. If the data of n-multiple variables could be plotted in a n-dimensional space, whose dimensions were scaled by variables' measures which defined the dimensions, the resulting plot would be a diffuse cluster plot of variable data points. Multivariate analysis that is dedicated to deriving orthogonal

---

*Note - The similarity between the results of the two multivariate analyses was predictable when considering the small number of variables factored during each analysis (Rummel,(46), page 519). Since the computer programs implemented during this study permitted the analysis of only 20 variables at a time, the close agreement between the two techniques was not surprising.

or independent factors, might be considered to be a mathematical technique that determines "factors" (hypervariables) which account for a large proportion of the variance of the entire data cluster. For example, Rummel(46) describes the principal-axis form of multivariate analysis as the technique utilized for determining the principal axis of the multidimensional ellipse of two or more of the multivariate dimensions. The axis defined through multivariate analysis are the minimum dimensions required to linearly reproduce the original data variance.*

Containing the geometric conceptualization, principal component analysis orthogonally rotates the coordinate axis of the multidimensional data cluster plot such that the original variables are now described by new principal components. The new components account for a maximum amount of variance of the data under consideration. The first principal component derived is the linear combination of the original variables which contribute the most in accounting for the total data variance. The second derived principal component is uncorrelated, orthogonal, and independent to the first component, and accounts for

---

*Note - Before proceeding further with the discussion, it must be noted that the tests by Rummel(46) and Harman(22) present an excellent discussion of the theory and techniques of multivariate analysis. Not too surprisingly, these texts are referenced a great deal in the following discussions.

the residual variance. The third component, and so on, accounts for the remaining independent residual data variance. The sum of the variances of the principal components should total to the sum of variance within the data. This method is optimal when the data is described in the same units of measure.

There are distinct differences between factor analysis and principal component analysis. Principal component analysis considers the total variance of the data defined by the variables. Factor analysis, though similar to principal component analysis in that it is a multivariate analysis, is distinct in that only the communality (a measure of the common variance of the data defined by the variables*) between the variables' data is analyzed as opposed to total variance. On the large data sets, the differences between factor analysis and principal component analysis become quite noticeable; i.e., the number of factors identified and the number of actual loadings will be noticeably different. However, in the case of this study, the small set of variables analyzed during each computer run gave rise to similar results.

---

*Note - For a more detailed understanding of the effects of error, Rummel's text gives an excellent explanation concerning data sets and the various errors that can be identified within a data set. Such errors are ultimately reflected in the variance of the data.

247

During the course of this study, the factor analysis of the data was performed using the library routines of the Syracuse University. The principal component analysis was performed using the "FACTAN" computer library routine of the RADC Information Processing Facility.

Data Set Variables

This study's multivariate analysis was primarily directed to investigating the inter-relationships and significance of the weather variables and the three categories of image-extracted variables with respect to the yield measurements' variables. The latter yield measurements included the mean yield of the three sites and the measured yield of the sixteen locations within the boundaries of the three sites. There were a few studies which investigated the relationships between weather data and ground truth data. However, these studies were minimized because the objectives were to investigate yield prediction using remotely-sensed data.

The image-extracted data variables which were considered during this multivariate analysis study are described in detail in the section entitled, "Digital Data Set" (pages 180 - 188), illustrated in Figure 64 (page 187), and rationalized in the section defining the crop yield prediction theory of this study, pages 25-39. The image-extracted data of the variables GLTX, YSAX, and DIFX were analyzed using the multivariate techniques of Factor and Principal Component Analysis.

248

When the particulars of the significant multivariate analyses are presented, the specific variables which were considered in the analysis will be specified. However, at this point of the discussion, it is assumed that preceding sections of this report presented adequate detailed information concerning the nature of the three categories of the data set.

Data Slice

According to Rummel(46), data may be described by three descriptors: entity, characteristic, and occasion. With respect to this feasibility study in crop yield prediction, Rummel's "entity" corresponded to a phenomena or event occurring at a location within the three sites' boundaries. Specifically, each entity corresponded to a minifield location where yield and other measures were collected. The latter measures were the "characteristic" descriptors defined by Rummel. Finally, each of the different measures or characteristics were collected at the "entities" (minifields) during different "occasions" of the growth cycle. Though the phenomena acting within a site location is defined by a three dimensional descriptor system (place-"entity"; measure-"characteristic"; and time-"occasion"), the multivariate analysis only investigated the relationships of two of these dimensions; therefore, the term "Data Slice" for this section.

The multivariate analysis was performed many times on various data sets so that specific relationships could be extracted. For example, a principal component analysis was performed to investigate each month's image-extracted variables and weather variables. This analysis assisted in the determination of which variables accounted for the greatest proportion of data variance on a month to month basis. Similarly, all of the image variables' data for the six months of coverage were factor analyzed. This analysis resulted in the determination of which spectral images and categories of image variables accounted significantly for the total data variance integrated over time. Though none of these analyses would permit conclusions about a variable's ability to predict yield, the analysis conveyed an idea of the variable's significance with respect to the other variables of the same class, etc. Table 16 lists the general information concerning the various principal component analyses which were performed during the course of this study.

Criteria for Data Exclusion from the Analysis

There were two basic reasons for certain variable's data being excluded from the principal component analyses. First, since the computer library program, FACTAN, permitted the analysis of only twenty variables at a time, only subsets of

250

Title	Variable Categories Analyzed	Objective of Analysis
Green	YSAX, DIFX, & GLTX green image data for April thru October	Determine if best components correspond to a specific image variable or to a particular month as a function of spectral image data utilized.
Red	YSAX, DIFX, & GLTX red image data for April thru October	Same as above objective
Near-Infrared	YSAX, DIFX, & GLTX near-infrared image data for April thru October	Same as above objective
New May New June New July New August New September New October	May's weather and image variables (all spectral images' data) → October's weather and image variables.	Determine if significant components consisted of image data variables or weather data variables as a function of the month. Same as above objective.
YSAX & Weather	YSAX data for all spectral images of April thru October and two weather variables' data for June and July.	Analysis may hint at whether the image data or weather data accounts for the most data set variance.
DIFX & Weather	DIFX data for all spectral images of April thru October and two weather variables' data for June and July.	Same as above objective.

Table 16. General information concerning some of the Principal Component Analyses performed to determine basic inter-relationships within the crop yield prediction data.

the eighty variables' data could be analyzed during a given computer run. The *second reason* for ommitting data was due to the occasional inability of the airborne multispectral camera to collect imagery over a site.

Limitations of the computer library program resulted in several computation runs using only subsets of the eighty variable data set. For example, perhaps all of the image-extracted variables and weather variables' data would be incorporated into a multivariate analysis which considered the data on an independent monthly basis. Another approach might be to factor analyze the image-extracted variables' data for the April thru October time period along with the weather data from only the key months of the growth cycle; pollination stage, July.

The failure to collect the required image data during the growth cycle occurred several times due to weather, cloud cover, pilot error, etc. For example, during July, a small area of Site Three was shadowed due to cloud cover at the time the multispectral photography was collected over the site (Reference: Figure 57, Page 177). This area occupied the location of one of the minifields where yield was measured (Reference: Ground Truth, pages 70,95). Consequently, the data from that area of the imagery could not be used; which prevented the analysis of all sixteen minifields' image-extracted data.

252

Ultimately, for other reasons, during certain months or series
of months, only 11 to 15 of the 16 minifields image data could
be incorporated into the principal component analysis. Table 17
presents a breakdown of the useable minifields' data for the
three sites as a function of specific months.

Time Period	Number of Minifields with Useable Data
May	15
May, July 2nd	15
May thru July 16th	14
May thru August	14
May thru September	13
May thru October	11

Table 17. Listing of the number of useable minifields' image-
extracted data as a function of time in the growth cycle.
Useable data could be used in the multivariate and regression
analyses.

Assessment of Error in the Data Set

The different methods of multivariate analysis are dis-
tinguishable by the manner in which each technique considers
error in the data set. For example, principal component
analysis, by its nature, considers the total variance of a

253

data set to consist of both systematic error and random error variance in addition to the common variance of the data set (Rummel, 46); Factor analysis considers only the common variance of the data.

Certainly, this data set, utilized for investigating the feasibility of crop yield prediction, must have numerous sources of error in each variable's measurements. For example, error in the form of false density information characterizes the digitized imagery to a degree due to problems in photo-graphic processing, exposure, digitization, etc. (Reference the report sections concerning digitization and the theory for crop yield prediction, pages 155 and 189 , respectively). Moreover, the weather data has some error in it also. The three sites were located 5 to 15 miles from the different weather stations where the weather data was acquired. Cer-tainly, the weather recorded at a station cannot necessarily be identical to that of a site. Despite the error, it is believed that the different weather variables' data and image-extracted data are appreciably accurate for defining the phenomena occuring in the field. In a future large scale feasibility study, better control of the data acquisition would be required. In the case of this study though, it can only be stated that the best error controls available were utilized.

254

Since the purpose of this program's multivariate analysis was to determine only the basic relationships of the variables prior to performing a regression analysis, the problems of accounting for the systematic or random error of the variables were ignored; thus, principal component analysis was emphasized instead of principal factor analysis.

Distribution Transformations and Matrix Transformations

Typically, when data is analyzed in a factor or component analysis, the raw data cannot be used directly due to the different scales of measure by which the data was acquired. To cope with this problem, data is usually preprocessed by the investigator such that a normalized data set is developed. The transformations could be linear, exponential, or any other transformation which would result in the normalization of the scaling effects.

Harman(22) contends that a multivariate analysis investigates a normalized data set when the correlation matrix of the factored variables is the prime source of data. As a consequence, some scientists calculate the correlation matrix from the raw data and input this matrix as the data for the multivariate analysis. Considering that this particular study was only a subsidiary effort preceding the regression analysis, this investigator chose to use the correlation of the raw data

255

as the prime transformation for producing a normalized data set which would be incorporated into each principal component analysis.

The Number of Factors or Components Criteria

When performing a multivariate analysis using factor or principal component analysis, determining how many of the factors or components are significant becomes a relevant question. To answer this question, both subjective and objective criteria can be used. In this particular study, which utilized the computer library program FACTAN, the objective criteria for terminating the number of identified significant factors or components occurred when the factor loadings (normalized eigenvectors of the variables' correlation matrix) became less than 1.0. The latter approach is distinct from others which might determine the number of significant factors based upon the amount of variance that the factors or components accounted for, scree tests, or statistical tests (Rummel, 46). Since there is a problem with defining the subjective cut-off point for determining the number of significant components, it is believed that the objective approach is optimal.

Rotation Technique

The FACTAN computer library program (described in the next section) utilized the Varimax Criteria for orthogonally rotating the principal components determined in the first phase

256

of the analysis. Geometrically speaking, the principal com-
ponent's axis, which are computed in the first phase of the
analysis, are rotated so that the new axis positions optimally
account for the variance of the data clustering in the n-
dimensional space. Rummel describes the Varimax criteria to
be dependent upon the variance of the factor loadings of a
given component. The criterion for performing the orthogonal
rotation is a function of the maximum variance calculated when
the component's squared component loadings (variance of the
component) is maximized (Reference: Rummel(46), pages 391 to
393).

FACTAN Principal Component Analysis

FACTAN was the principal computer library program which
was utilized to perform the principal component analysis of
the crop yield prediction program's data. FACTAN is documented
by Honeywell Corporation(25) and is implemented on the General
Computer Operating System(GCOS) of the Rome Air Development
Center's Information Processing Facility. FACTAN performs a
principal component analysis of multivariate data and determines
the optimal "components" which describe the general dimen-
sionality of the total data variance. This section will
describe the mechanics of FACTAN in moderate detail.

Initially, FACTAN calculates the correlation matrix from
the raw data of the multiple variables. After analyzing the

257

intercorrelation within a set of variables, the significant component loadings are computed by component analysis and rotations are performed via the varimax rotation criteria. The ouptut of FACTAN can be interpreted for determining the minimum number of independent dimensions needed to account for most of the variance of the original data set.

Specifically, FACTAN performs the following sequence of steps: "N" sets of observations from "M" sites are input to the program; FACTAN calculates the M x M correlation matrix where the diagonal elements are defined as 1.0 (criteria for a principal component analysis) which results in the total variance being accounted for rather than just the common variance); component loadings, $V_i = 1, \ldots, M$, or the normalized eigenvectors of the correlation matrix are computed; components' loadings with eigenvalues L_i less than 1.0 are considered significant and the variance accounted for by the factor loading V_i is proportioned to L_i; and finally, the initial component loadings are rotated to obtain the structure of the data (principal components) which best account for the total data variance (Reference: Honeywell(25)).

The actual computer dialogue of FACTAN is printed in Appendix D. The results of the various principal component analyses are reported in Appendix E.

Interpretation of the Multivariate Analyses

Introduction

Referencing Appendix E, this section presents the inter-
pretations of some of the significant principal component
analyses which were performed with the image and weather data.
Appendix E summarizes most of the principal component analyses
which were performed with the FACTAN computer program. A list
of the variables, which were considered in the multivariate
analyses, with the variables' respective definitions is located
at the beginning of Appendix E. Generally, each of Appendix
E's analysis summaries will include a title, the list of the
variables considered, notes, a comment describing the objective
of the particular study, and a review of the important results
of the analysis.

Image Variable's Significance per Spectral Band

The first three investigations of the multivariate analysis
study sought to identify the relative significance of the three
image-extracted variables, YSAX, DIFX, and GLTX, as a function
of spectral band. This objective was satisfied by performing
three separate principal component analyses (Reference:
Appendix E, page 432). During each analysis, the three
image variables' data which was acquired from a particular band

259

of April thru October imagery, was incorporated into the FACTAN data set. Each FACTAN-computed principal component accounted for a proportion of the data set's variance. The component's highly loaded variables corresponded to a particular image variable type. For example, the green band analysis, Appendix E (page: 435), resulted in the determination of five principal components. The first component, which accounted for 38% of the total variance of the data set, primarily consisted of the variable GLTG of the mid-summer months (Reference: Variable Definitions, page 433). This predominance was determined by comparing the factor loadings of GLTG to those of the other variables incorporated into the first principal component. Similarly, the second component, which is both orthogonal and independent to the other components is composed primarily of the GLTG variable of the early and late months of the growth cycle.

By referencing the results of the green, red, and near-infrared band analyses on pages 434-36 of Appendix E, some conclusions can be made about the relative significance of the image variables. The most significant image data type was identified with the image variable GLTX, which is the mean multispectral density of the total pixels occupying the imaged site. Furthermore, the green and red spectral images' GLTX

data accounted for a large proportion of the data set variance earlier in the growth cycle (July 2nd thru mid-August) than the near-infrared GLTX data which predominated during the late months of the growth cycle. Finally, the second general component was identified to be some combination of the DIFX and YSAX variables and does not seem to be characterized by any particular spectral image's data.

Significance of Image versus Weather Data per Month

Principal component analyses were performed on each month's weather and image-extracted variables' data. Each month's data analysis was performed independent of the other months' data. This multivariate analysis study permitted the determination of the importance of weather data relative to image-extracted data in accounting for the total variance of the subject data set. Appendix E, pages 437 , are the summaries of each month's analysis.

After all six component analyses' results were interpreted, it was concluded that the weather data and GLTX data defined the first principal component and the DIFX data made up the second component. The YSAX variable was found to be relatively unimportant in accounting for the total variance of the subject data set as compared to the weather, GLTX, and DIFX data.

261

Furthermore, this analysis indicated the independency of the DIFX data versus the GLTX data and a high correlation between weather data and the GLTX image data. The latter relationship has a utility in predicting the yield of imaged areas defined by the YSAX and DIFX variables. Since weather is an important predictor variable of yield, the high correlation of weather with GLTX variables suggests that the two are interchangeable. This would mean that the yield of a crop could be predicted by utilizing the information extracted from variables such as GLTX and DIFX. This apparent relationship is investigated in more detail during the course of the regression analysis study.

Finally, these analyses indicated that the green and red spectral images' information that is incorporated into the DIFG and DIFR variables respectively, accounted for more variance than the DIFIR variable. This is substantiated by the predominance of the DIFG and DIFR variables being incorporated into the second component as compared to the DIFIR variable, which was included in the third orthogonal and independent component.

YSAX & Weather Variables/DIFX & Weather Variables

The two variables that defined the density and density residual of the imaged corn minifields of each site corresponded to the YSAX and DIFX variables respectively. The YSAX

variable described the mean density level of the 72 pixel areas within the digital spectral images of the site. The DIFX variable described the difference between the YSAX value and the mean density level of the entire imaged site (GLTX).

The section of the report concerning the theory of crop yield prediction, page 189 , makes reference to the bias against the information value of the YSAX variables' data due to the lack of calibration in the data set. The same section also states that the rationale for creating the DIFX variable was to empirically correct the problems normally associated with the YSAX data. The two principal component analyses which considered YSAX & weather variables and DIFX & weather variables respectively, was an attempt to determine the significance of the image variable relative to the weather data from a key period of the growth cycle, (Reference: Table 16, page 251). Though these two analyses are not documented in Appendix E, they are reviewed in this section because they are referenced later in the discussion of the regression analysis.

There are two reasons for considering only the weather data from June and July. The first reason for the limitation was due to the fact that only 20 variables could be analyzed by FACTAN at a time. Secondly, the weather data from June and July was utilized because the literature concerning corn

263

production emphasized the importance of the weather during the post-germination thru pollination stages of the corn growth cycle. If any weather data was to be considered, that of June and July seemed to be most appropriate.

The first principal component analysis investigated the YSAX and weather variables. The data consisted of all the YSAX image-extracted data for the months of April thru October and the weather data described in five variables' data for June and July. The prime component was identified with the YSAR and YSIR data from the July 2nd imagery. The secondary component was attributed to the June weather data. Finally, the third component was associated with the YSAG data from May. The results suggest that the June-July period may be a dynamic period of the growth cycle. Weather was identified to be secondary to image data for the total variance of the data set.

The second principal component analysis considered the DIFX data, which had been extracted from the April thru October multispectral imagery, and the weather data of June and July. The DIFX-weather data analysis computed a principal component which consisted of the variable DIFG and DIFR for the image collection periods of July 2 and July 16. The second component consisted of the DIFG variable of May and the DIFIR variable of July 16. This analysis indicated that the weather

264

data of an important period of the growth cycle was not significant in accounting for data set variance until the third component was calculated.

Interpretation of the Correlation-Matrices

During the course of the multivariate analysis, the image-extracted variables' data and the weather variables' data were input to a subroutine which calculated the correlation matrix for later implementation in the principal component analysis. Tables 18 to 21 summarize the important results from this analysis.

Table 18 summarizes the inter-correlation between the image-extracted variables; that is, the correlation between GLTX/DIFX, GLTX/YSAX, and DIFX/YSAX. Table 18 may be interpreted to mean that the mean density level of the multispectral images of the large   field (GLTX) is independent (low correlation) of the DIFX data which describes the density differential between selected stressed areas with the large field (YSAX) and that of the large field (GLTX). If a regression model equation required independent data for predicting yield, the GLTX/DIFX combination would be preferable to the other combinations.

Monthly Image Variable Correlation Pairs	Low Correlation $C < .6$	Minor Correlation $.6 \leq C < .8$	Moderate Correlation $.8 \leq C < .9$	High Correlation $C \geq .9$
GLTX/DIFX	M, J_2, J_{16}, A, S, O			
GLTX/YSAX		J_{16}	J_2	A, O
DIFX/YSAX		J_2, A, O	J_{16}	M, S

Table 18. Intercorrelation Matrix of Image Variables

Table 19 summarizes the intercorrelation of the image variables' spectral components. Since each image variable is composed of three spectral images' data, the image variable can actually be broken down into three multispectral component variables; i.e., GLTX - GLTG, GLTR, and GLTIR. The correlations used for developing Table 19 indicate that digital multispectral data is highly intercorrelated. Such a fact is indicated by the high intercorrelation of the variables dependent upon actual grey level, specifically, GLTX and YSAX. The amount of intercorrelation was reduced somewhat in the case of the DIFX variable; correlation was only a moderate, $0.8 \leq C < 0.9$. The latter observations are important when considering that a future regression equation for predicting

266

Monthly-Image Inter-Correlation Pairs	Low Correlation $C < .6$	Minor Correlation $.6 \leq C < .8$	Moderate Correlation $.8 \leq C < .9$	High Correlation $C \geq .9$
GLTX/GLTX	M	S		J_2, J_{16}, A, O
DIFX/DIFX		A	M, J_2, J_{16}, O	S
YSAX/YSAX		M, S		J_2, J_{16}, A, O

Table 19.   Intercorrelation of the image variables with variables of the same category but different spectral band (i.e., GLTG correlated with the variables GLTR or GLTIR).

yield would require predictive variables which are minimally intercorrelated for optimal prediction accuracies. This means that to use highly intercorrelated variables in a regression equation reduces the amount of distinct information being input for predictive purposes.

Table 20 illustrates some very important relationships. Since the original corn minifields were selected according to different stress conditions within the corn field, the imaged minifields' density fluctuation is a function of localized stress and not due to weather. This accounts for both the low correlations between the YSAX or DIFX data with weather data. However, Table 20 also indicated a high correlation between

Monthly Image/Weather Correlation	Low Correlation $C < .6$	Minor Correlation $.6 \leq C < .8$	Moderate Correlation $.8 \leq C < .9$	High Correlation $C \geq .9$
YSAX/WEATHER	M, S	J_2, J_{16}, A, O		
DIFX/WEATHER	M, J_2, J_{16}, A, S		O	
GLTX/WEATHER			O	M, J_2, J_{16}, A, S

Table 20. Correlation between image data and weather data.

GLTX data and weather data. Assuming that the weather conditions are generally uniform over a given site and believing the fact that a measure of a crop's photosynthetic potential is the plants' canopy albedo, weather data and the average albedo of a crop should be correlated (Reference: Literature Review of Crop Yield Prediction Studies, pages 25-39).

A significant inference evolves from these correlation results between the GLTX and weather variables. If weather data is a key predictor of yield, then the possibility of replacing weather data with image data (i.e., GLTX) for yield prediction purposes may be feasible. This hypothesis is investigated further in the regression analysis study.

Table 21 presents the correlation of the image-extracted variables with weather as a function of spectral band. Again this analysis supports the data of Table 20 which stated that

| | Green | | | |
Spectral Image/Weather Correlation	Low Correlation $C < .6$	Minor Correlation $.6 \leq C < .8$	Moderate Correlation $.8 \leq C < .9$	High Correlation $C > .9$
YSAG/WEATHER	M, S	J_2, J_{16}, A, O		
DIFG/WEATHER	M, J_2, J_{16}, A, S		O	
GLTG/WEATHER		M	A, O	J_2, J_{16}, S
	Red			
YSAR/WEATHER	M, J_2, S, O	J_{16}, A		
DIFR/WEATHER	M, J_2, J_{16}, A, S, O			
GLTR/WEATHER	O	J_2, A, S	J_{16}	M
	Infrared			
YSIR/WEATHER	M, J_2, J_{16}, S, O	A		
DIFIR/WEATHER	M, J_2, J_{16}, A, S, O			
GLTIR/WEATHER	S, O	J_{16}		M, J_2, A

Table 21. Correlation of Image Data with Weather Data as a function of spectral band.

of the three variables, GLTX correlated highly with weather.
Table 21 substantiates Table 20 and adds the inference that
the principal bands for weather correlation were the green
and infrared; specifically, the GLTG and GLTIR variables.

Conclusions

The multivariate analyses served as a prelude to the
multiple regression analyses. Significant conclusions cannot
be expressed at this time. This decision is attributable to
the caution with which the investigator regards this analysis.
One reason for the caution is due to the fact that the com-
ponent analyses were limited by both the number of variables
that could be analyzed and the useable data of the data set.
Nevertheless, the relationships that were observed provided a
guideline about which the regression analysis was structured.
Finally, the relationships which were concluded from the
component analysis did support many of the relationships
identified through the analysis of the correlation matrices.
One of the conclusions that were inferred from the interpre-
tation of the multivariate analyses was the fact that certain
image variables' data could account for more data set variance
than the weather data. Furthermore, of the three image
variables, GLTX data demonstrated an ability to account for
more of the data set variance than the YSAX or DIFX data.

However, neither of the two latter image variables could be judged as superior to the other. A better comparison test could be conducted using regression analysis techniques. The multivariate analyses also indicated that the imagery collected during the image collection periods of July 2 and July 16 accounted for a good proportion of data set variance. By identifying which variables accounted for most of the variance of the data set, the data set variability could be described by those specific variables. Practically speaking, instead of predicting yield with many image or weather variables, yield could be predicted using only a few.

Though the multivariate analysis did not indicate which variables were best for predicting yield, they did indicate which variables were independent and accounted for a significant portion of the data set variance. Such information may be useful when variables are selected for implementation into a yield prediction regression model which requires both independent and significant variables' data.

271

Multiple Linear Regression Analysis for Crop Yield Prediction

Introduction

Laboratory evidence neither supports or disproves the
hypothesis that a linear relationship exists between an agri-
cultural crop's albedo and its yield. For a crop yield pre-
diction feasibility study seeking first-order relationships
between image data and crop yield data, a linear relationship
can be assumed. Subsequently, linear regression analyses were
employed to identify the basic relationships between a crop's
yield and its multispectral image data and/or weather data.*

Prior to performing the multiple linear regression analysis
studies, the ground truth, weather, and image data had been
preprocessed utilizing the multivariate analysis techniques
of principal component analysis and factor analysis. However,
with the exception of the inferences derived from the analysis
of the correlation matrices, no study had identified a relation-
ship between a crop's yield data and corresponding remotely-
sensed data. This section of the report describes the linear

---

*Note: During the course of this discussion, the author assumes
that the reader is knowledgeable of the jargon and theory of
multiple linear regression analysis. The text by Draper and
Smith(15) is an excellent reference which clearly describes
both the theory and applications of regression analysis.

272

regression analysis studies which were performed in behalf of investigating the feasibility of crop yield prediction using remotely-sensed data.

Research Question

The literature that is pertinent to crop yield prediction indicates that a relationship exists between the crop yield and specific ground truth information; such as, row spacing, plant population, ground moisture, etc. For example, if the presence of fertilizer in a field could be incorporated into a yield prediction model; e.g., by utilizing numbers which measure the degree of farm management practices within the field, the yield of the field could be partially predicted. However, since the intent of this study was to investigate the relationship between remotely-sensed data (i.e.: multispectral imagery of a crop) and crop yield, the regression analysis de-emphasized the analysis of ground truth data with respect to crop yield. Instead, the regression analyses investigated the possible relationships between weather and/or image data to yield. The research question of the regression analyses considered the following areas of inquiry:

1) The relationship between crop yield and each image variable's temporally-variant data (i.e., YSAX and DIFX) as a function of spectral band (i.e., green, red, near-infrared);

273

2) The relationship between crop yield and each image variable's three spectral components (i.e., YSAX-YSAG, YSAR, & YSIR) as a function of time;

3) The yield prediction potential of the variables identified as significant in the principal components analyses;

4) The relationship between crop yield and each image variable's data extracted from combinations of two spectral images (i.e., YSAG & YSAR; YSAG & YSIR; or YSAR & YSIR) as a function of time;

5) The relationship between crop yield and both weather data and image data which was extracted from combinations of two multispectral bands;

6) The relationship between crop yield and each month's image and weather variables' data;

7) The relationship between crop yield and key predictive months' image-extracted data and weather data.

In summary, the research question of this study was directed to determining the relationship between yield data and spectrally-controlled data, temporally-controlled data, weather data, and combinations of the latter categories.

Data Slice of the Regression Analysis

During the course of the data acquisition effort, three categories of data had been collected; namely, image data,

ground truth data, and weather data. Certainly, a few regres-
sion analyses were performed which linked the ground truth
data to yield (i.e., row spacing, barren plants, etc.). How-
ever, due to the reasons cited in the section defining the
research question, the ground truth data was de-emphasized
from consideration. Consequently, the weather and image-
extracted data were primarily considered during this regression
study.

The total remotely-sensed data set was comprised of six
months' data. Each month's data consisted of five weather
variables and three basic image variables (i.e., YSAX, DIFX,
and GLTX). Each image variable could be subdivided into three
spectral component variables (i.e., GLTX - GLTG, GLTR, and
GLTIR; Reference: Variable Definitions, Appendix G, page 451).
Since the computer library program for multiple linear regression
analysis could only analyze 24 variables at a time, the entire
data set could not be analyzed. However, by analyzing subsets
of the entire data set (data slices), multiple studies were
performed which investigated various relationships within the
data set. Consequently, many of the interpretations of the
various regression analyses illustrate the relationship of yield

275

as a function of spectral information, spectral band com-
binations, weather, or weather-imagery variable combinations.*

Regression Analysis Techniques Utilized

The techniques of stepwise regression (SWR), forced-forward regression (FFR), and forced-reverse regression (FRR) were utilized in satisfying the objectives set forth in the research question section. The stepwise regression analysis was useful in determining which variables were significant predictors of crop yield. The two types of forced regression analysis were valuable in determining which months' data was significant in predicting the yield of crops.

The stepwise and forced regressions were performed with the computer library program entitled "SMLRP" which is imple-mented on the RADC Information Processing Facility's general purpose computers. Following a computation of the data variables' correlation matrix, SMLRP calculates a sequence of multiple linear regression equations in a stepwise manner.

---

*Note: Appendix G lists the various regression analyses per-formed during the course of this investigation along with the general variables considered in each study. The various data sets utilized in the regression analyses were often handi-capped by the number of minifields' data which could be utilized during the particular study. The reasons for reject-ing data were described previously in a section of the report describing the multivariate analysis of the data, page 252 .
For a better understanding of the data set, the multivariate analysis section is recommended reading.

During each step, only one variable is added to the developing linear regression equation. Each variable that is added corresponds to the variable which produces the greatest reduction in the error of the sum of the squares. This could be restated to mean that each added variable has the largest F-value. Such variables have the highest partial correlation to the dependent variable, crop yield, with respect to the variables already incorporated into the equation. Each additional variable has a diminishing, but optimal, ability to account for the variability of the dependent variable.

SMLRP also provided the capability of performing both forced-forward regression analysis and forced-reverse regression analysis. These regression techniques permitted the forced inclusion of variables in the crop yield prediction regression equation. This approach contrasts with the stepwise regression approach which included the most significant predictor variables into the equation.

The forced regressions were invaluable in determining the significance of adding consecutive months' image and weather data into a yield prediction equation. The significant months for yield prediction were also identified utilizing these techniques and the results supported the conclusions derived from the stepwise regression analysis. Furthermore, these forced

277

regressions indicated that the predictive power of key variables was diminished if too much data was utilized in the regression analyses for yield prediction.

By definition, the forced regressions were performed at a low level of confidence. However, the stepwise regression analyses produced results at both a 90% confidence level and a low level of confidence defined by F=0.01. The 90% confidence level permitted conclusions to be drawn from the regression analyses which were considered to be statistically valid with a high level of confidence. The low confidence level permitted the stepwise regressions to incorporate all of the variables into the equation such that trends in the selection of yield predictive variables could be studied.

Technical Description of the SMLRP Computer Library Program

The Stepwise Multiple Linear Regression Program (SMLRP), which is documented by Honeywell(25) and implemented on the RADC Information Processing Facility's General Computer Operating System(GCOS), performed the stepwise multiple linear regressions and forced regressions of the image data and weather data. The Fortran program computes by steps the least squares optimal value for the coefficients of an equation of the form:

$$Y = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + \ldots B_n X_n$$

where: Y is the dependent variable, yield (bushels/acre)

$X_1, X_2, \ldots X_n$ are the independent variables (image variables: GLTX, YSAX, or DIFX; weather variables)

$B_0, B_1, B_2, \ldots B_n$ are the coefficients of the linear equation which are to be computed by the SMLRP program.

Some of the features of the SMLRP program are: the program handles up to 24 variables; stepwise addition and deletion of independent and dependent variables at run time; variable transformations of input data; ability to force variables into regression; residual summary available; option to enter critical F-value or confidence level; user provided variable labels.

SMLRP performs a stepwise regression analysis using the matrix of correlation coefficients. One or more independent variables and one dependent variable may be selected from either the input or transformed variables. In the stepwise procedure, the independent variables are added one at a time to the regression to give the intermediate equations:

$$Y = B_0 + B_1 X_1$$
$$Y = B_0' + B_1' X_1 + B_2' X_2$$
$$Y = B_0'' + B_1'' X_1 + B_2'' X_2 + B_3'' X_3$$

279

Statistical information is obtainable in each step of the computations. At each step, those independent variables not included in the regression are inspected to determine which variable provides the greatest reduction in the variation of the dependent variable, Y. This variable is then tested for significance; i.e., the computed F-ratio of the variable is compared to the supplied critical F-value. If the computed value is greater than the critical value, the variable is considered significant and is added to the regression solution.

After each variable is added, those variables already incorporated in the regression are inspected to determine if any of them can now be deleted because their contribution to the reduction in the variance of the dependent variable is no longer significant. Those variables which now have a computed F-ratio less than the critical value are considered insignificant and are deleted from the regression solution. This process is continued until no more variables can be added or deleted. Thus, the final regression solution contains only those variables that are statistically significant.

It is possible to force one or more of the independent variables into the regression solution. Without regard for the statistical significance, such variables are added to

the linear fit before any other variables. The remaining

independent variables are then added to the fit in the normal

stepwise procedure as reviewed in the preceding paragraphs.

However, once a variable has been forced into the regression,

it will not be deleted in the following steps even though it

is not statistically significant.*

The dialogue of the stepwise regression computer library

program, SMLRP, is presented in Appendix F, pages 445 - 449

of this report.

Interpretation of the Regression Analysis

Introduction

Approximately twenty regression analysis studies utilized

the SMLRP computer library program. These studies analyzed

image data, weather data, and yield data so that both the

inter-relationships of the data variables and the feasibility

of crop yield prediction using remotely-sensed data could be

determined. Rather than attempt to develop the ideal "yield

prediction equation" during this feasibility study, the inves-

tigation was limited to the interpretation of the regression

analyses' results. The results typically included the amount

---

*Note: The preceding paragraphs paraphrase the technical
description of the SMLRP program documented by Honeywell(25),
pages ST-145,146.

of yield data variance, R^2, that could be accounted for by the regressed independent variables' data as a function of time; the order in which the variables were selected for incorporation into an equation developed by stepwise regression analysis; the type of variables which were selected for incorporation into a regression equation that predicted yield* with a high degree of confidence, etc.

The results of each of the nineteen regression analysis studies were graphically plotted (i.e., Reference Figures 97 thru 102, etc.) and interpreted. Only the significant relationships and conclusions extracted from the interpretation of the graphed regression analyses' results are discussed in the following sections. Conclusions derived from interpretations of small scale phenomena observed in the plotted results were abandoned. The latter decision was made after considering the scope of this feasibility study and the statistical significance which could be assigned to the results. Conclusions based on small scale observations were considered premature and

---

*Note: Regression analyses which utilized data acquired early in the growth cycle and which demonstrated an excellent ability to account for the variance of the minifields' yield data, were considered to have effectively simulated the prediction of the yield of these minifields. The independent variables of the latter regression analyses were considered to be yield predictors and the regression equation was considered to have demonstrated yield prediction.

exemplified fallacious reasoning. Consequently, most inter-
pretations' conclusions are conservative in nature. However,
it will be discovered that even these conservative conclusions
are significant and may be revolutionary if they prove to be
valid in future studies.

Appendix G, pages 450-57, presents a tabular summary of
the important regression analysis studies performed using the
SMLRP program. A list of the variables implemented in the
regression analyses, along with the respective variables'
definitions, is incorporated at the beginning of Appendix G.
Each of the Appendix G's regression analysis summaries includes
a title, the objective of the study, the types of regression
analyses performed, the variables considered, the time period
in the growth cycle during which the variables' data was
acquired, and a reference to the figure which graphically
presents the results of the regression analysis.

283

Interpretation of Regressions #1 thru #6 (Ref: Appendix G)

The first six regression analysis studies evaluated the
predictability of agricultural crop yield and compared the
yield predicting potential of the YSAX and DIFX image
variables as a function of spectral band, utilizing temporally-
registered spectral image data. In other words, these studies
analyzed the temporally-registered green, red, and near-
infrared image data of the variables YSAX and DIFX independently
in order to evaluate each image variable's ability to predict
the yield of the sixteen minifields.*

Figures 97 thru 102 graphically illustrate the results of
the first six regression studies. The vertical scale of each
figure measures the amount of yield data variance,R^2, which
could be accounted for by the regression analysis of the image
variable's data. The horizontal scale plots the six occasions
of the corn growth cycle during which useable image data was
collected. Each figure contains three multicolored curves.
Two of the curves record the results of two regression analyses.
The third curve plots the absolute numeric difference between
the two regressions' plotted results.

---

*Note: The discussion of the regression analysis results will
only make sense if a clear understanding of the data set and
theory supporting this feasibility study is established. The
reader should familiarize himself with the information of the
previous sections describing the data set and variables, espe-
cially pages 180-199 , before proceeding with this discussion.

284
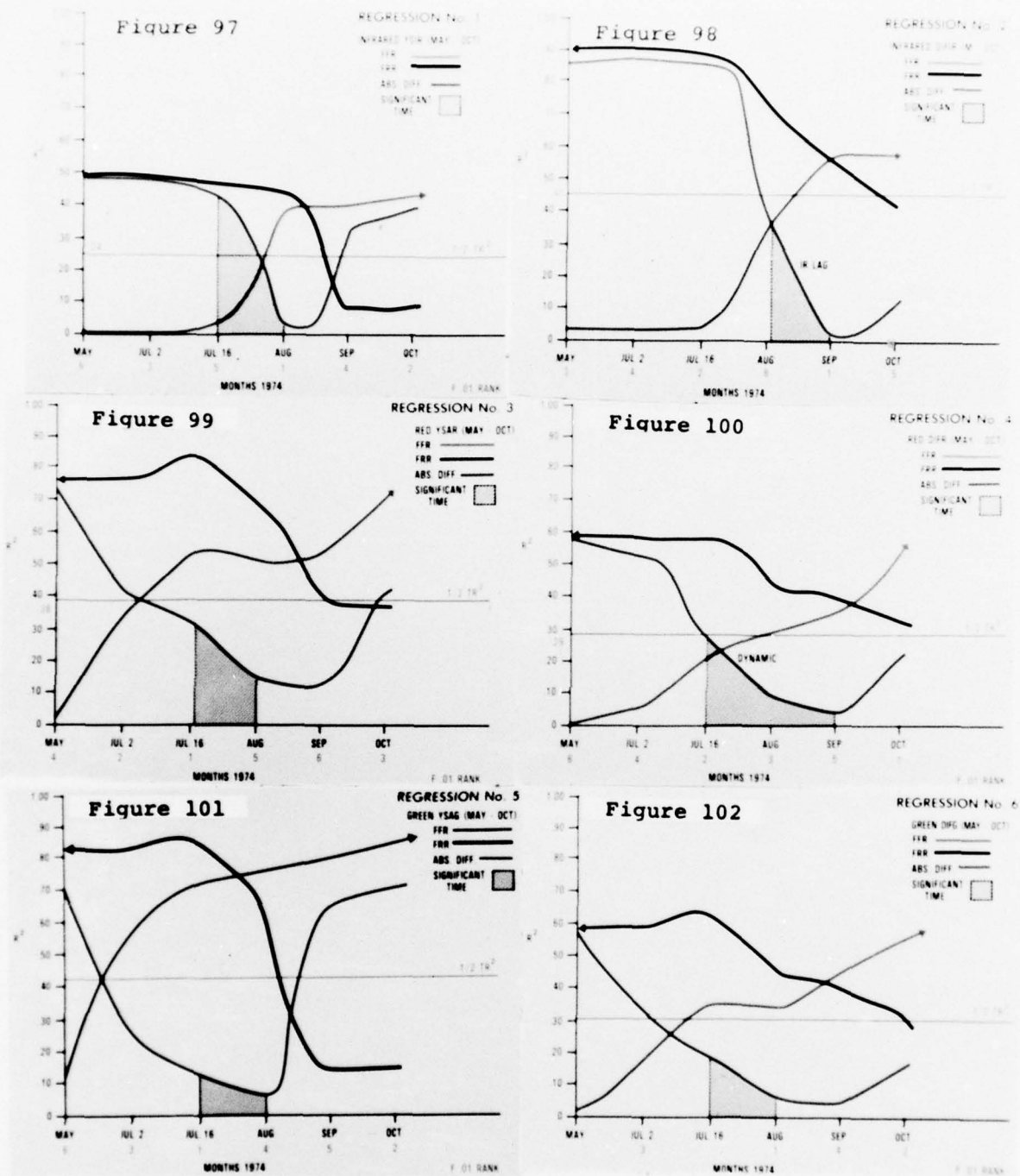
Figures 97 thru 102: Regressions 1,3,5 and 2,4,6 illustrate the utility of
each separate multispectral image's independent contribution to predicting
yield using YSAX and DIFX image-extracted variables respectively(Ref: Figure
64, page 187). The graphs indicate that an ideal period for yield prediction
may occur during the mid-July to mid-August period. R^2 is both a measure of
the total yield data variance accounted for by the regression-analyzed image
variables and of how well crop yield can be predicted. FFR, Forced Forward
Regression, and FRR, Forced Reverse Regression, were two types of regressions
performed on the data. The titles imply that either May-October or October-
May data were forced into the yield prediction regression equation.

285

Three regression analyses were performed during each of the six studies; specifically, a forced-forward regression (FFR); a forced-reverse regression (FRR); and a stepwise regression (SWR). The data set of each regression study consisted of the YSAX or DIFX spectral image variables' data (i.e., YSAG, YSAR, DIFG, DIFIR, etc.), extracted from the May thru October temporally-registered images, and the final yield data (YSA) of the 11 minifields which had excellent multispectral image data during the latter time period.

The forced-forward regression analysis (FFR), which consisted of a series of six regressions, was accomplished in the following manner. In the first regression, only May's image data was regressed with the yield data. For example, in the case of the first regression analysis study, Figure 97, the regression of May's data yielded poor results. The yield data's variance could not be accounted for (predicted) by May's image data (Ref: green curve of the FFR of May's image data). In the second regression step, the image data collected during the July 2nd period was added to May's data set and the total data set was regressed with yield data. Again, in the case of Regression Study #1, Figure 97, the yield data's variance could not be predicted. However, as the process was repeated by adding July 16th image data, followed by the image data of

286

August thru October, the amount of yield data variance accounted for increased. Figure 97 indicates that 40% of the yield data variance could be accounted for using the YSIR data for the months of May thru August.

The forced-reverse regressions (FRR) were performed in a similar fashion as the forced-forward regressions, except that each month's image data was incorporated into the sequential regressions in reverse order. This means that October's image data was regressed with yield data initially. Subsequent regressions added the image data from the months of September, August, etc. The blue curve of Figures 97 thru 102 represent the results of the FRR analyses.

The third graph in each of the six figures is a red curve labelled "absolute difference". This curve was a plot of the absolute numeric difference of the yield data variance which could be accounted for by the FFR and FRR analyses of the image data. This graph was useful in qualitatively determining the best periods of the growth cycle during which the subject image data was most useful in predicting yield.

Below the horizontal axis of each of the six figures are small red numbers. These numbers indicate the order in which the third regression analysis of the data set, a stepwise regression, selected the best data for yield prediction. For example,

287

Figure 97 indicates that the YSIR data of August was considered most significant in predicting the yield data variance of the 11 minifields; followed by the image data of October, July 2, September, July 16, and May. Though the stepwise regression analysis was performed at a low confidence level, (F-critical = 0.01), it did indicate the ranked order by which the variable's monthly data was selected for predicting yield. It is premature to state which months are best for yield prediction; however, it is useful to note each of the six studies' results now for later reference.

A measure of a good yield predictor variable was considered to be the amount of yield data variance that the variable could account for in a regression analysis. By comparing Figure 97 to Figure 98, some conclusions could be made concerning the predictive ability of the YSIR variable versus the DIFIR variable. Figures 97 and 98 illustrate that the YSIR and DIFIR variables demonstrated similar yield predicting abilities using the May thru July 16th image data. However, over a period of six months, DIFIR appeared to be a better predictor of yield than YSIR. This observation was also derived by interpreting the level of the $1/2 \ TR^2$ line in the two figures, which plots the level of one-half of the total yield data variance accounted for during the entire FFR analysis.

288

If yield prediction were feasible using temporally-registered image data, knowledge of optimal periods for imagery collection would be desirable. It would be more timely to predict yield using imagery collected during the months of, say, June and July rather than using image data collected from May thru August, etc. Consequently, part of the interpretation of the six regression studies was devoted to extracting any suggestion of optimal image data collection periods for crop yield prediction.

Since laboratory experiments had identified that a crop's albedo was a measure of a crop's health, dynamic changes in a crop's health could be indicated by the change in the crop's albedo (Reference: Literature Review of Crop Yield Prediction Studies, pages 25-39). Figures 97 thru 102 indicate that a significant change in a crop's multispectral albedo does exist. By utilizing the "absolute difference" curve, the dynamic periods of change in the crop's multispectral albedo were easily interpreted. The identified periods' corresponding image data was later re-evaluated thru regression analysis techniques to determine the feasibility of predicting yield using image data collected during short periods of the growth cycle.

A subjective criteria was utilized in determining the dynamic periods of the growth cycle which would be optimal

image collection periods for achieving yield predictions. The criteria required that the absolute difference curves have a high negative slope preceding a zonal minimum. For example, Figure 97 indicates that an important period exists between July 16th and August. Besides matching the criteria, this time period demonstrated the greatest increase in the amount of yield data variance that was accounted for in the successive FFR analyses. Similarly in the case of Figure 98, the DIFR variable's regression indicates that the August thru September period, and perhaps the July 16th thru September period, may be an ideal imaging period for yield prediction. Figures 97 thru 102 have blackened regions beneath the "absolute difference" curves. These regions mark the time intervals that were subjectively determined to be optimal periods predicting yield using spectral image data.

Several conclusions were derived from the interpretation of the six figures concerning the relationship between image data and crop yield. For example, the optimal period for yield prediction utilizing the multispectral image data of the crop, described by the variable YSAX, was the July 16th thru mid-August period (for all spectral bands).

The DIFX variable seemed to require the image data from September before it could significantly account for the yield data variance.

290

The YSAX spectral image data was observed to account for more yield data variance than DIFX image data earlier in the growth cycle, with the exception of the near-infrared data.

The green and red spectral image data of both the YSAX and DIFX variables were better able to predict yield than the near-infrared image data of the two variables. Near-infrared image data demonstrated a noticeable lag time (approximately one month) in its ability to predict yield. This fact was demonstrated by the high yield data variance that could be accounted for by the green and red image data as early as July 16th in comparison to that accounted for by the near-infrared image data.

The aforementioned conclusions are interesting in a couple of ways. For example, the optimal period for collecting imagery for yield prediction, which corresponded to the dynamic period of change in the crop's multispectral albedo, also corresponded to the pollination and post-pollination stage of the growth cycle. This period has been long identified as a key period of the growth cycle by both farmers and agronomists. Secondly, with the exception of the near-infrared image data, the DIFX image data demonstrated a poor ability, compared to the YSAX image data, to account for the yield data variance. This is surprising when it is realized that the YSAX image data is

virtually raw image density data which was subjected to numerous processes which allegedly degraded the data's information content. It is ironic that the DIFX variable, which was created to empirically remove the image degrading effects of atmosphere, exposure, film processing, and digitization, demonstrates such a poor ability as a potential yield predictor. However, it is essential to keep in mind that the first six regression analysis studies investigated spectral image data independently. Therefore, it may be premature to conclude anything absolute about the yield prediction capability of a particular image variable until both image and weather data are analyzed.

Interpretation of Regressions #7 and #8 (Ref: Appendix G)

The seventh and eighth regression analysis studies evaluated the predictability of agricultural crop yield and compared the yield predicting potential of the YSAX and DIFX variables, utilizing all three spectral bands of temporally-registered image data. In other words, these studies analyzed the temporally-registered green, red, and near-infrared image data of the variables YSAX and DIFX so that each image data set's ability to predict yield could be evaluated.

Figure 103 and 104 graphically illustrate the results of the seventh and eighth regression analysis studies. The vertical and horizontal scales are defined similarly to the figures interpreted in the previous section. Each of the figures contains six curves corresponding to three regression analyses performed on two different data sets. The two data sets were analyzed by using the techniques of forced-forward regression (FFR); stepwise regression (SWR) at a 90% confidence level; and a stepwise regression at a low confidence level (F-critical = 0.01).

Two different data sets were analyzed during each study in order to optimize the information that could be extracted from the temporally-registered multispectral image data. The first data set consisted of the image and yield data of the

293

Figure 103

Figure 104

Figures 103 & 104: Regressions 7 & 8 illustrate the utility of using all three multispectral images' data for predicting yield using the YSAX and DIFX image-extracted variables respectively(Ref: Figure 64, page 187). Beginning with May's image data, each month's image data was consecutively added to the regression data set to determine how early in the growth cycle could yield be predicted.

Definitions:

MAX & MIN OBS: Sample size used in the regression analysis-indicates the maximum number of corn fields per month which had complete image coverage from May to the subject month. Only 11 fields had coverage from May to October. Note: Separate regressions were performed for both MAX & MIN sample sets.

FFR: Forced Forward Regression(See pages 276 - 281 for explanation).

F.01: Stepwise regression at a minimum confidence(accuracy)level: useful for determining which variables were significant for yield prediction

0.90CL: Stepwise regression whose R^2 value gives a measure of how much crop yield is predictable with a 90% accuracy.

R^2: Total yield data variance that could be accounted for in the regression analysis

eleven minifields, located within the boundaries of the three agricultural test sites, which had excellent multispectral image data during the months of May thru October. This data set was entitled "MIN OBS" in each of the figures. The second data set's dimensions varied as a function of time and corresponded to the maximum number of minifields, located within the boundaries of the three sites, which had excellent multispectral image data during the specified period of time. The number of minifields in the maximum observation category was annotated beneath the horizontal scale of each figure. This data set was entitled "MAX OBS" in each figure. For example, Figure 103 indicates that 15 minifield's image data collected during May, and yield data were analyzed. However, when the image data of May thru July 16th were analyzed, only 14 minifields data (MAX OBS) could be analyzed. Similarly, for the May thru October imagery, only the image data from 11 minifields and corresponding yield data could be considered in the regression analysis of the "MAX OBS" data set. The reasons for this variability in useable data was explained in previous sections which described the data set. It will suffice to say that the minifields that were excluded from consideration in the regression analyses were omitted because no imagery was collected over the minifields during that particular time for various reasons.

295

The procedures utilized during these two regression analysis studies were similar to those of the previously interpreted regression analyses For example, Figure 103 illustrates the results of the regression analysis study which investigated the relationship of the green, red, and near-infrared image data of the variable YSAX with the yield data of the useable minifields. Each series of regressions was accomplished in the following manner. In the first regression, of say, the forced-forward regression analysis (FFR) of May's image data, the green, red, and near-infrared YSAX image data was regressed with the yield data of 11 to 15 minifields (Reference the green or black curve for May). In the second step of the series of the FFR analyses, the green, red, and near-infrared YSAX image data collected during May and July 2nd were regressed with the yield data of the 11 to 15 minifields. This process was repeated for each additional month's image data. The results, in terms of the amount of yield data variance accounted for, R^2, are plotted as a function of the months whose image data was considered.

Each of the regression analysis techniques was useful in extracting information from the data set under consideration. The stepwise regressions (SWR) at the 90% confidence level were useful in determining the best yield predicting image data.

Furthermore, the stepwise regressions performed with a low level of confidence (F-critical = 0.01) typically rank ordered the data set's image variables as a function of their ability to account for the yield data variance. Finally, the forced-forward regressions indicated the importance of adding temporally-registered image data to a regression equation; specifically, it increased the amount of yield data variance accounted for and demonstrated that not all data was useful for yield prediction - in fact, too much can be detrimental.

The index in the upper right hand corner of Figures 103 and 104 illustrate which curves correspond to a particular series of regressions of the maximum and minimum data sets. The code "FFR" represents the results of the forced-forward regressions. The codes ".90C" and "F.01" represent the stepwise regressions which produced results at the two extremes in accuracies or confidence levels.

Figure 104 graphically presents the results of the regression analyses of the green, red, and near-infrared DIFX image data and yield data. The procedures and graphic representations are identical to those previously described for the seventh regression study. Interpretation of Figures 103 and 104 provide a comparison of the multiband YSAX and DIFX image variables as predictor variables of crop yield.

297

Several important relationships were derived from the interpretation of the two regression studies' results. Probably the most significant regression analyses were those performed at the 90% confidence level which utilized the data from eleven minifields. For example, the stepwise regressions of the YSAX and DIFX data, which were performed at the 90% confidence level (red-dashed curves of Figures 103 & 104), indicated that 85% of the yield data variance could be predicted as early as August with a 90% accuracy. In the case of the analysis of the YSAX image data, Figure 103, the results were even more remarkable. The SWR analysis at the 90% confidence level indicated that as early as July 16th, 87% of the yield data variance was predictable with a 90% accuracy.

When comparing the results presented in Figures 103 and 104 to those of the first six regressions, it can be qualitatively stated that the ability to predict yield using remotely sensed data increases with the number of spectral bands of image data. In the first six regression studies, which independently considered a single band of spectral image data for yield prediction, the amount of yield data variance that could be accounted for seldom exceeded $0.70R^2$ during the forced-forward regressions. The FFR analyses, of Figures 103 and 104, accounted for more yield data variance in the early months than did the FFR analyses

298

of Figures 97 thru 102; the only distinction was that three bands of image data were utilized instead of a single band of image data (later regression studies utilizing two band combinations of YSAX or DIFX image data also bear this conclusion out).

The results of these regression analyses of the YSAX and DIFX image data indicate that the YSAX variable is a more powerful predictor of yield than the DIFX variable. This is concluded by the fact that for all six curves which plot the results of six regression analyses, the YSAX image data analyses accounted for a significantly greater proportion of the yield data variance earlier in the growth cycle than did the analyses of the DIFX data. However, considering the theory upon which the YSAX and DIFX variables were respectively created (Ref: pages 180 - 199), it might be premature to assume that the YSAX variable was superior to the DIFX variable as a yield predictor variable.

Two other features were observed in the graphs of Figures 103 and 104 that were considered noteworthy. The regression analyses of the YSAX May thru July 16th image data indicated a steady increase in the amount of yield data variance that could be accounted for. However, all of these curves indicated that the amount of yield data variance that was accounted for

299

diminished during the July 16th thru August period and finally increased at the end of the growth cycle. The plotted results of the DIFX data analyses are quite different in that they indicate a great proportion of yield data variance could be accounted for during the July 16th thru August period with a diminishing predictive ability after August. The reason for this observed phenomena may become apparent during later analyses.

Finally, the results of the regression analyses of the "MIN OBS" data set demonstrated a superior ability to account for yield data variance than the analysis of the "MAX OBS" data set. This was thought to be attributed to the variability in the dimensions of the "MAX OBS" data set as a function of time; the "MIN OBS" data set was always constant. Since the statisticians place more trust in results generated by a larger data set, the results of these two studies should be considered with guarded optimism.

300

Interpretation of Regressions #11 and #12 (Ref: Appendix G)

The eleventh and twelfth regression analysis studies evaluated the predictability of agricultural crop yield and compared the yield predicting potential of the YSAX and DIFX image variables, utilizing two band combinations of temporally-registered spectral image data. In other words, these studies analyzed two band combinations of the temporally-registered green, red, and near-infrared image data sets (i.e., green/red, green/near-infrared, etc.) of the variables YSAX and DIFX such that each combination's ability to predict yield could be evaluated.

Figures 105 and 106 illustrate the results of the eleventh and twelfth regression analysis studies of the YSAX and DIFX image data respectively. The vertical and horizontal scales are defined similarly to the figures interpreted in the previous sections. Each figure contains three curves corresponding to the forced-forward regression analysis (FFR) performed on each of the three image data sets. The image data collected during May thru October and the yield data of eleven corn minifields were considered in the regression analyses. Though the results of only the forced-forward regressions are graphed in Figures 105 and 106, a stepwise regression analysis at the 90% confidence level was also performed.

301

In comparing the results of the FFR analyses illustrated in Figures 105 and 106, the YSAX image data demonstrated a superior ability to account for the variability of the yield data variance than did the DIFX image data. More yield data variance could be accounted for in the May thru October period (especially the May thru July 16th period) by the YSAX image data than the DIFX image data.

Figure 105

Figure 106
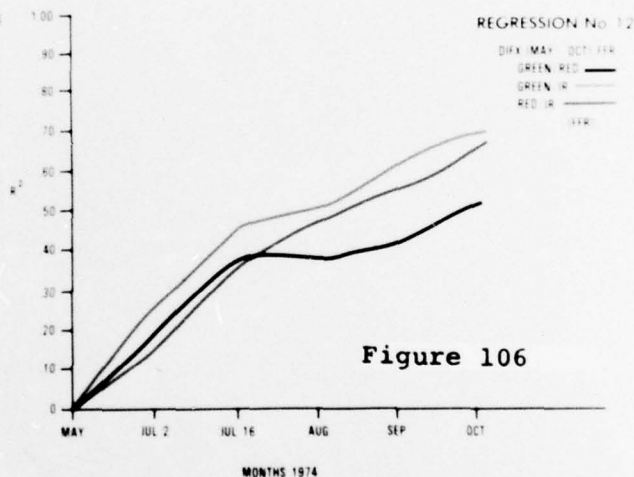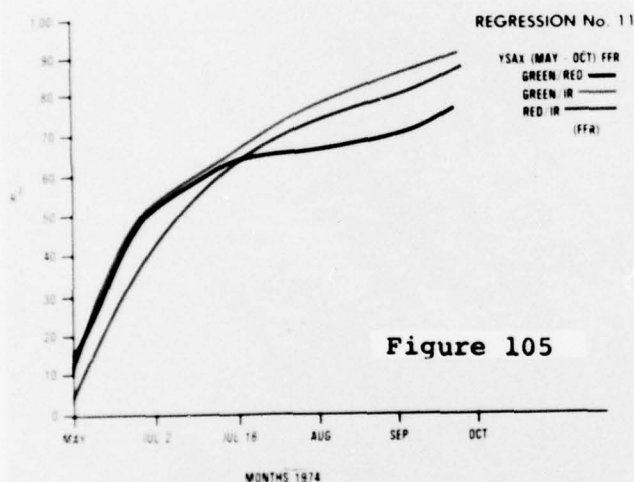
Figures 105 & 106. Regressions 11 & 12 illustrate the advantages of certain two band combinations of multispectral image data for predicting crop yield. Graphs plot the results of a forced forward regression.

302

According to the FFR analyses, the best two band combinations of spectral image data for yield prediction was the green/near-infrared image combination for both the YSAX and DIFX variables respectively.* The green/red image data accounted for a more significant proportion of the yield data variance during the early months of the growth cycle than the red/near-infrared combination. However, during the later months of the growth cycle the converse of the latter statement was true.

The amount of yield data variance accounted for by the FFR analyses of the two band combinations of the image data exceeded that predicted by the regression analysis of the single band image data. This conclusion was derived through simple qualitative comparisons of Figures 105 and 106 with Figures 97 thru 102. Furthermore, the regression analysis of the three bands of image data (green and black curves of Figures 103 and 104) did not account for a significantly greater proportion of yield data variance than did the regression of the two band combinations of the YSAX and DIFX image data. However, in the analysis of the same data using

---

*Note: The stepwise regression analyses indicated that the red/near-infrared data was the best image data combination for yield prediction. The results were generated at the 90% confidence level.

303

stepwise regression analysis techniques, three bands of image data had a greater utility for predicting yield than did either the two band or single band combinations of image data. Certainly though, the FFR analyses of the three band combinations of image data, Figures 103 and 104, indicate how too much data can provide suboptimal yield predicting capability.

The stepwise regression analysis (SWR) of the two band combinations of image data resulted in the determination of variables which could account for 87% of the yield data variance with a 90% accuracy. These outstanding yield predictor variables consisted of the two band combinations of the YSAR/YSIR data (contradictory of the results of the FFR analyses). Similar combinations of the DIFR/DIFIR image data (Reference: Variable Definitions, Appendix G) could only account for 45% of the total yield data variance. In general though, the SWR analyses indicated that all combinations of the YSAX variables were better yield predictors than any two band combinations of the DIFX image data.

304

Interpretation of Regressions #13 & #14 (Ref: Appendix G)

The thirteenth and fourteenth regression analysis studies evaluated the predictability of agricultural crop yield and compared the yield predicting potential of the YSAX and DIFX image variables, utilizing the two band combinations of temporally-registered image data and weather data. These regression analysis studies were similar to the two preceding studies (Ref: Figures 105 & 106), except that monthly weather data was considered along with the image data. Furthermore, these studies utilized stepwise regression analysis techniques during the course of the investigations.

Figures 107 and 108 illustrate the results of the regression analyses of the YSAX/weather and DIFX/weather data respectively. The vertical and horizontal scales are defined similarly as those of the figures interpreted in the previous sections. Each figure contains three curves corresponding to the stepwise regression analysis (SWR) which was performed on each of the three data sets. The analyzed data set consisted of the image data and weather data collected during May thru August and the yield data of fourteen corn minifields. The SWR analyses were performed at the low confidence level

305
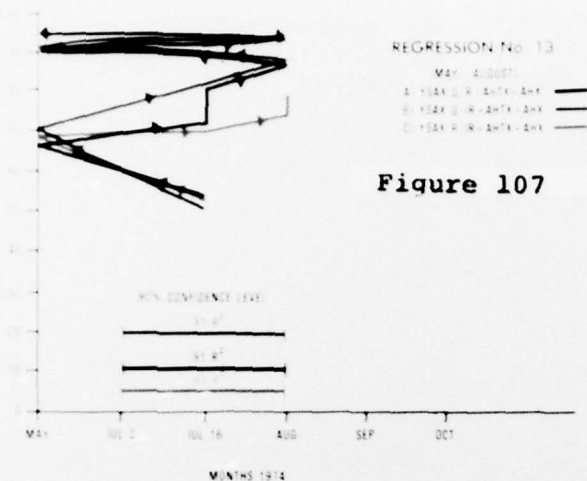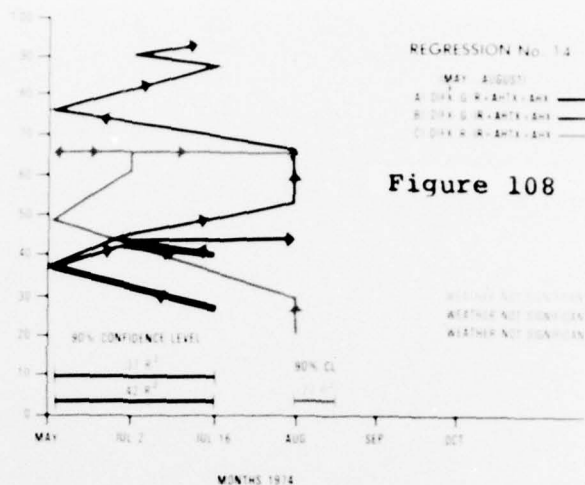
Figures 107 & 108. Stepwise Regressions 13 & 14 illustrate the advantages of certain two band combinations of multispectral film data when regressed with certain weather data for yield prediction. The arrows on the graphs reflect the month's data which was considered most significant for yield prediction. YSAX variable was a better predictor than DIFX. Optimal dates for yield prediction span from 2 July through mid-August. Weather was significant when used in conjunction with the YSAX variable.

(F-critical =0.01). However, the results were interpreted at the 90% confidence level so that image or weather variables, which were identified as significant yield predictors, could be identified.

Since the literature review of crop yield prediction research and the results of this feasibility study's multi-variate analysis indicated the significance of weather data in predicting crop yield, weather data was incorporated into

306

the regression analysis of the two band combinations of the YSAX and DIFX image data. After interpreting the results of the multivariate studies of the weather, which was analyzed in conjunction with image data, the variables defining the average high temperature per month (AHTX) and the average humidity per month (AHX), were incorporated into the SWR analyses. Only the image and weather data of the months of May thru August were selected for study because this time period was subjectively judged to be optimal for demonstrating yield prediction. If incorporating September and October's data were required, the results of such regression analyses would be considered to be yield measurement rather than yield prediction.

The procedure for performing the two regression analysis studies was somewhat different than that of the previously interpreted regression studies. In this case, there were no series of regressions which incorporated each additional month's data into the equation. Instead, the subject image data of the months of May thru August was included, along with the two types of weather data from the same months, in the stepwise regression analysis (SWR). The SWR analysis was performed at a low confidence level such that all of the data's variables were eventually selected as being significant for yield prediction.

Figures 107 and 108 graphically illustrate which months'
data were selected for inclusion into the regression equation
in the order in which the selected data was determined to have
diminishing significance in predicting crop yield. For example,
Figure 107, which records the regression results of the YSAX/
weather data, indicates that July 16th's data (image or weather)
was selected as the most significant data for predicting yield
data variance of the 14 minifields. The vertical scale
indicates that the data (image or weather) of July 16th
accounted for approximately 55% of the yield data variance.
Furthermore, the three curves indicated that May's data (image
or weather) was considered to be of secondary significance in
predicting yield. In fact, when combining the May data with
the July 16th data, a total of 70% of the yield data variance,
R^2, could be accounted for. Similar logic may be used to
interpret the remaining portions of the graphs.

Figures 107 and 108 are extremely valuable for determining
which two band combination of image data, in conjunction with
weather data, was most significant for yield prediction.
For example, in Figure 107, the red curve terminates after
accounting for only 75% of the yield data variance; whereas,
the other two curves extend themselves until nearly 95% of
the yield data variance is accounted for. Though this result
offers hints as to which data set is best for yield prediction,

308

a better measure exists in each of the figures. These measures are the horizontal multicolored lines which are parallel and adjacent to the horizontal axis of the figure. The horizontal lines describe the time interval in which the corresponding image data and weather data (Ref: Color Code Key in each figure) could account for a certain proportion of yield data variance with a 90% accuracy (90% level of confidence). For example, in Figure 107, the horizontal green curve may be interpreted to mean that both the green/red YSAX image data combination and weather data acquired from July 2nd thru August (19th) predicted 91% of the yield data variance with a 90% accuracy.

From the latter introduction to the interpretation of Figures 107 and 108, it can be recognized that these two studies' results are extremely useful in determining the best of the two potential yield predictor variables (YSAX or DIFX), the best two band combinations of image data, and the optimal time periods for yield prediction using image and weather data.

During the analysis of the YSAX image and weather data, Figure 107, weather variables were never identified as being very significant for yield prediction purposes. However, weather variables were occasionally incorporated into the equations involving the YSAX image data. In the case of

Figure 108, which plots the results of the stepwise regression analysis of the DIFX image data and weather data, weather data was considered insignificant. This was demonstrated by the fact that weather data was never included in the regression equation, even at the low confidence level.

Figure 108 clearly illustrates the inferiority of the DIFX variable as a yield predictor. In the best case in which the DIFG and DIFR data were regressed with weather data, only 42% of the yield data variance could be accounted for at a 90% confidence level. Furthermore, in comparing Figures 107 and 108, the DIFX/weather data never came close to matching the ability of the YSAX/weather data in accounting for the yield data variance. After interpreting the results of the two studies at the 90% confidence level, it is apparent that the optimal two band image data combinations, when considered with weather data, were the green/red and green/near-infrared combinations. Moreover, the optimal period for predicting yield with a 90% accuracy, using the YSAX/weather data, spanned from early July thru mid-August.*

Finally, weather data was found to be a relatively insignificant predictor of yield compared to either of the two

_____

*Note: Though the DIFX/weather data analyses indicated an earlier period for yield prediction, the early months' data could only account for 42% of the yield data variance.

310

image variables, YSAX or DIFX. Between the two image variables though, the YSAX and weather data were optimal combinations for crop yield prediction.

Despite the fact that the two analyses indicate the inferiority of the DIFX variable as a predictor of yield, this investigator is reluctant to state that any yield prediction should be a function of raw image data (i.e., YSAX data) rather than a residual density measure (i.e., DIFX data). This reluctance is based on a knowledge of the characteristics of the image data sets. For example, though every effort was made to control the photochemistry process, digitization, etc., there were inherent image degradations which ultimately affected the true information content of this image data set. Certainly, any operational system would be required to generate calibrated imagery for many reasons which ultimately relate to the computer's ability to semiautomatically identify crops and predict yields. This feasibility study included the development of the image variable DIFX in order to empirically remove the effects of image degradation. Unfortunately, for unknown reasons, the DIFX image data is demonstrating a poor ability to account for the variance of the yield data.

Before abandoning the idea of utilizing an image's density residual or density variance as a measure for crop yield prediction, another study should be performed to re-evaluate

311

the utility of certain types of image data. Most importantly though, future studies investigating crop yield prediction should utilize multispectral imagery which is calibrated (i.e., electro-optical scanner sensors' imagery).

Interpretation of Regression #15 (Ref: Appendix G)

The fifteenth regression analysis study evaluated the predictability of agricultural crop yield as a function of each month's image and weather data. By considering each month's data separately in a regression analysis for yield prediction, another means of determining which months' data was most significant for yield prediction could be investigated. The following regressions were performed:

May - 29 April imagery with May's weather data

June - 2 July imagery with June's weather data

July - 16 July imagery with July's weather data

August - 19 August imagery with August's weather data

September - 4 September imagery with September's
            weather data

October - 17 October imagery with October's weather data.

Figure 109 illustrates the results of the fifteenth regression analysis study. The vertical and horizontal scales are similarly defined as those of the figures interpreted in the previous sections. The figure contains three color-coded "saturns" which correspond to the three different types of regression analyses which were performed on each month's image and weather data sets. The image data (i.e., GLTX, DIFX, and YSAX), collected during each month, and the yield data of eleven minifields were included in each month's data set

313

which was considered by the regression analyses. Only the
minifields which had excellent image data during all six months
were utilized in this study. This precaution insured that no
particular month could demonstrate a superior yield predicting
capability due to its results being based upon a different data
set or sample size.
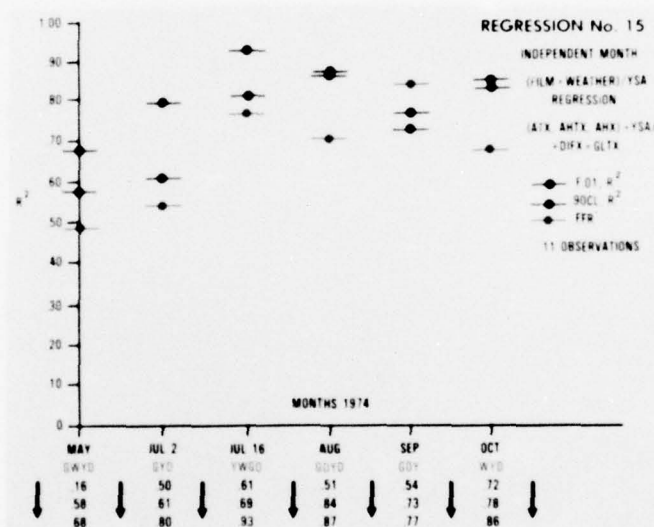
Figure 109: Regression 15 illustrates which month's image and
weather data are most significant for yield prediction. Three
regressions were performed for each individual month's data:
Stepwise regressions at both low and high accuracies (F.01
and 0.90CL, respectively); Forced forward regression. July 16
and August data are most significant for yield prediction. For
variable definitions reference Figures 103 & 104, page 294).
Special Definitions: ATX-Average Temperature/month; AHX-Average
Humidity/month; AHTX-Average High Temperature/month.

314

Each month's data set was analyzed by three types of regression analysis techniques; specifically, a forced-forward regression (FFR); a stepwise regression (SWR) at a 90% confidence level; and a stepwise regression analysis at a low level of confidence (F-critical = 0.01). Referencing the color code index in the upper right corner of the figure, each regression analysis's results were coded in color (i.e.: FFR - red saturn; SWR @ 90%CL - green saturn; and SWR @ F.01 - blue saturn).

Before proceeding with the interpretation, there is one other unique characteristic of Figure 109 which requires discussion. At the bottom of the figure, below the horizontal axis are a horizontal series of alphanumeric characters consisting of the letters G, W, Y, or D. Below the letters is a vertical column of numbers which increase in magnitude as they are read downwards. This additional information presents the results of the SWR analysis performed at the low confidence level. The intent of the symbols and numbers is to illustrate the order and significance of the variables extracted from the data set for inclusion into the regression equation. For example, beneath the July 16th image collection period are the letters "YWGD" and the numbers "0.61 thru 0.93". This means that the order in which the variables were selected for inclusion in the SWR equation was the following: YSAX ("Y"),

315

which accounted for 61% of the yield data variance; Weather ("W"), which accounted for an increase of only 8%, totalling 69%, of the yield data variance; and GLTX ("G") and DIFX ("D") variables, which together accounted for another 24% of the yield data variance. Continuing the example, the green saturn, marking the results of the July 16th SWR analysis at the 90% confidence level, indicates that 80% of the yield data variance could be accounted for with a high degree of confidence, using this data set. This latter detailed information permits an interpreter to identify the significant month's significant data which was included in the yield prediction equation.

Interpretation of Figure 109 indicates that the image and weather data collected during the mid-July thru mid-August time interval demonstrated the best ability to account for the yield data variance. Though October's analysis indicated that October's image and weather data could significantly account for the yield data variance, October could not be considered a good yield prediction period because of the month's temporal relationship with the growth cycle.

Weather variables were found to be insignificant yield predictors during the months of June (July 2nd imagery collection), August and September. Remembering the fact that weather data and GLTX data were discovered to be highly

316

intercorrelated during the multivariate analysis, it seemed
that there was a good possibility that the GLTX variables
were being substituted for the weather variables in the yield
prediction regression equations.

Finally, it was observed that the FFR analyses illustrated
that the yield data's variance could not be "predicted"
optimally when the entire data set of a given month was incor-
porated into a regression equation in a forced fashion. The
exception to this observation occurred during September. Since
this was the only time that a FFR analysis demonstrated better
results than a SWR analysis, the results for September were
discounted as coincidental.

Interpretation of Regression #16 (Ref: Appendix G)

The sixteenth regression analysis study evaluated the predictability of agricultural crop yield utilizing image and weather data acquired during July 16th and August (19th). Since this time interval had been identified in past studies as being a significant period for acquiring image and weather data for yield prediction, several regression analysis studies were performed to investigate this period's significance more closely. This regression study initially analyzed specific image and weather data of July 16th. Upon completion of the July analysis, similar data from the August period were added to the July 16th data set and the analysis was repeated. The objective was to determine how much more yield data variance could be accounted for by combining and analyzing the data sets of the two periods.

Figure 110 illustrates the results of the sixteenth regression analysis study. The vertical scale is defined similarly to those of the figures which were interpreted in previous regression studies. The horizontal scale is abbreviated; that is, it includes two horizontal intervals for presenting the results of both the July and the July + August data analyses respectively. Figure 110 contains six vertical graphs which correspond to the three types of

318

regression analyses performed on the July and July + August
data sets. The image data collected during July 16th and
August (19th), the July weather data, and the yield data of
fourteen corn minifields comprised the general data set of
these regression analyses.
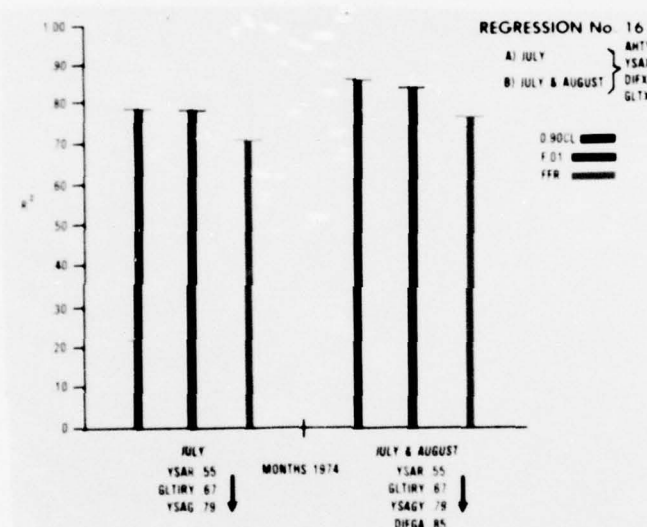
Figure 110 Regression 16 illustrates the significant increase
in yield data variance accounted for when July's image-extracted
data variables and single weather variable is combined with
similar August data in a regression equation for crop yield
prediction.

Each time period's data set was analyzed by three types of regressions; specifically, a forced-forward regression (FFR); a stepwise regression (SWR) at a low confidence level (F-critical = .01); and a stepwise regression (SWR) at a 90% confidence level. Referencing the color code index in the upper right corner of the figure, each regression analysis's plotted results were color coded (i.e.: SWR @ 90% CL - blue; SWR @ (F_c = 0.01) - green; and FFR - red).

The specific image and weather data set which was considered in the two series of regression analyses is listed in detail in the upper right corner of Figure 110. Since the average high temperature of July, AHTY, had been identified in both the multivariate analyses and the 15th regression study as a significant yield predictor variable, it was included in the regression analyses of the July and July + August image data sets. The average high temperature of August was not included in the analysis, because weather data from August was found to have insignificant value as a yield predictor variable (Ref: Regression #15). The image data set consisted of all the spectral components (i.e.: red, green, near-infrared) of the image variables GLTX, DIFX, and YSAX.

At the bottom of Figure 110, below the horizontal axis, is a column of variables. Adjacent to the variables is a

320

vertical column of numbers which increase in magnitude as they are read downward. This additional information presents the results of the stepwise regression analysis performed at the 90% confidence level. The intent of presenting this information was to illustrate the order and significance of the variables extracted from the data set for inclusion into the regression equation. For example, beneath the July interval on the horizontal scale are the variables and numbers: YSAR (0.55), GLTIRY (0.67), and YSAG (0.79); (Ref: Variable Definitions, Appendix G). This nomenclature means that when the SWR analysis at the 90% confidence level was performed, the latter variables were incorporated into the yield prediction equation. Each variable's data accounted for the corresponding percentage of yield data variance (i.e., July 16's YSAR image data accounted for 55% of the yield data variance).

Certainly, a yield prediction equation would not be ideal if it required the incorporation of several different image variables' data as indicated by the analysis of the July + August data sets. Future studies should ultimately determine which image data and months are truly useful for yield prediction.

Figure 110 illustrates that the SWR analysis, performed at the 90% confidence level, demonstrated that 85% of the

321

yield data variance could be accounted for when only the image and weather data of July 16th and August (19th) were analyzed. The results indicated a 6% increase in the yield data variance that could be predicted, occurred when both period's data were utilized. Though the increase was not overwhelming, the total yield data variance which could be accurately accounted for, by using image data acquired early in the growth cycle, was remarkable.

Though selected weather data, in the form of average high temperature data of July, was included in the regression analysis, the weather variable was found to be an insignificant yield predictor. This could be attributed to the fact that the weather data is truly insignificant or it could be due to the fact that weather information is accounted for by the GLTX image variable. GLTX data and weather data were found to be highly intercorrelated during the multivariate analyses.

It is encouraging to notice that the requirement for weather data may be nonexistent for yield prediction purposes, if a temporally-registered multispectral image data set is available. Though it may be fallacious reasoning to conclude at this time that weather data is not required for yield prediction, at least in the instance of this feasibility study, weather variables demonstrated negligible potential as yield predictors.

322

Interpretation of Regression #17 (Ref: Appendix G)

The seventeenth regression analysis study evaluated the predictability of agricultural crop yield utilizing July 16th and August (19th) temporally-registered image data and specific weather data of July. Since the results of the preceding study indicated that more yield data variance was "predicted" when the combined data of July 16th and August was regressed, the 17th regression study investigated which image data of the period was most useful in predicting yield of the fourteen minifields.

Figure 111 illustrates the results of the seventeenth regression analysis study. The vertical and horizontal scales are defined similarly to Figure 110, which was interpreted in the previous section. Figure 111 contains three vertical graphs which illustrate the regression analysis results of the three image/weather data sets. Each image/weather data set was analyzed by three regression analysis techniques; specifically, a forced-forward regression (FFR); a stepwise regression (SWR) at a low confidence level (F-critical = 0.01); and a stepwise regression (SWR) at a 90% confidence level. The amount of yield data variance that could be accounted for in each regression analysis of an image/weather data set is annotated on the multicolored graphs of Figures 111 as horizontal lines with labels (i.e., FFR-"FFR"; SWR @ 90% CL-".90CL"; and SWR @ $(F_c = 0.01)$ - "F.01").

323
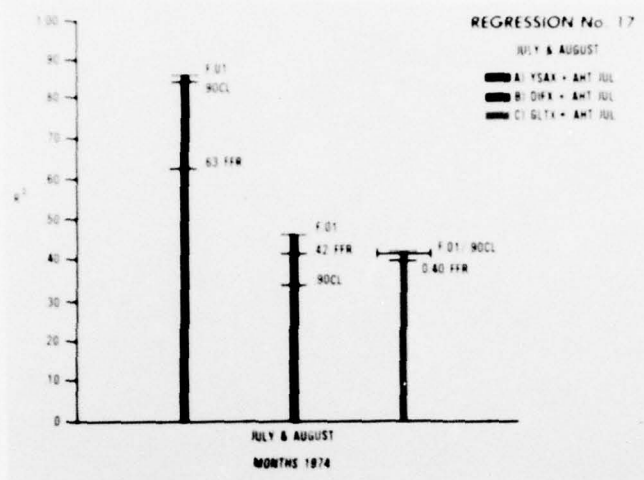
Figure 111: Regression 17 illustrates the amount of yield data variance accounted for when specific image/weather data combinations are employed in the regression analysis. Three types of regressions were performed: Stepwise (F.01), Stepwise (0.90CL or 90% accuracy); and forced forward regression (FFR), (See definitions, Figures 103 & 104).

The July 16th and August (19th) general data set consisted of the image data (i.e., GLTX, DIFX, & YSAX) and July weather data; specifically, the average high temperature of July, AHTY. As mentioned in preceding sections, the weather data, AHTY, has been consistently identified as significant data for yield prediction. However, when the GLTX data was incorporated into the regression equations, it seemed that even the significant July weather data became unimportant in predicting yield (Ref: 16th Regression Study).

324

In order to determine which image data was most significant in predicting yield, three data sets were considered: YSAX data from July 16th and August imagery with the AHTY weather data (green graph, Figure 111); DIFX data from July 16th and August imagery with the AHTY weather data (blue graph, Figure 111); and GLTX data from July 16th and August imagery with the AHTY weather data (red graph, Figure 111).

Figure 111 indicates that yield prediction is best achieved with the YSAX/weather data set of the July 16th - August (19th) period. The green graph illustrates that 85% of the total yield data variance could be described with a 90% accuracy using the YSAX/weather data set.

Though not apparent in the figure, AHTY weather data was included in the stepwise regression equation. In past studies, the weather data and GLTX data were determined to be extremely intercorrelated; and therefore, interchangeable. The next regression study will seek to determine if accurate yield prediction is feasible using YSAX image data from July 16th and August and the GLTX data from July 16th imagery collection period. By including both the GLTX and weather data in the regression analysis data set, the study will determine if the image variable GLTX is substituting for the weather variable in accounting for the yield data's variance.

325

Interpretation of Regression #18 (Ref: Appendix G)

The eighteenth regression analysis study evaluated the predictability of agricultural crop yield utilizing July 16th and August (19th) temporally-registered image data and specific weather data of July. This study was different from the preceding study in that two types of image data from July 16th and August were regressed with weather data from July. Since weather data and GLTX image data had been determined to be highly intercorrelated, the objective of this study was to determine if GLTX image data would be selected for inclusion into the yield prediction equation instead of weather data. The 17th regression study had identified the average high temperature of July, AHTY, to be a significant predictor of yield, when utilized in conjunction with the YSAX image data of both July 16th and August (19th). Both the AHTY weather data and the GLTX data of July 16th were regressed with YSAX data and DIFX data respectively during two separate analyses.

Figure 112 illustrates the results of the 18th regression analysis study. The vertical and horizontal scales are similarly defined as those of preceding figures. Figure 112 contains two vertical graphs which present the results of the regression analysis of the two image data sets. Each image data set was investigated using three regression analysis

326

techniques; specifically, a forced-forward regression (FFR);
a stepwise regression (SWR) at a low confidence level (F-
critical = 0.01); and a stepwise regression (SWR) at a 90%
confidence level. The amount of yield data variance that
could be accounted for in each regression analysis is
annotated on the multicolored graphs of Figure 112 as hori-
zontal lines with labels (i.e.: FFR-"FFR"; SWR @ 90% CL-
".90 CL"; and SWR @ (F_c = 0.01) - "F.01").
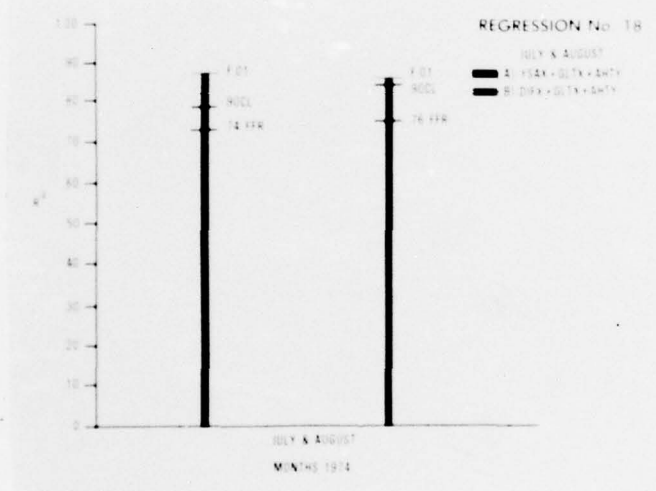
Figure 112: Regression 18 illustrates the amount of yield data
variance accounted for when specific image/weather data com-
binations are employed in the regression analysis. Three types
of regressions were performed: Stepwise (F.01), Stepwise (0.90
CL or 90% accuracy); and forced forward regression (FFR), (see
definitions, Figures 103 & 104).

This 18th regression study could be considered as a series of regression analyses of two data sets. The first study analyzed the YSAX and GLTX image data of July 16th and August and the AHTY weather data. This study's results were color coded blue in Figure 112. The second study analyzed the DIFX and GLTX image data of July 16th and August and the AHTY weather data. The latter study's results are color coded green in Figure 112.

This 18th regression study demonstrated that yield prediction of an agricultural crop is feasible utilizing only temporally-registered multispectral image data. The yield of 14 minifields was accurately predicted (90% confidence level-accuracy) early in the growth cycle; specifically, during the pollination and post-pollination phase of the growth cycle. Furthermore, when appropriate image data was available, weather variables were demonstrated to be insignificant yield predictors.

This particular study was unique in that its results did not support past observations which suggested that the DIFX variable was an insignificant predictor of crop yield. The interpretation indicates that the regression demonstrated timely and accurate crop yield prediction (90% accuracy) of

328

85% of the total yield data variance of 14 minifields utilizing a data set comprised of mid-July and August DIFX and GLTX image data.

The yield prediction equation, developed by the SWR analysis at the 90% confidence level, predicted the yield of fourteen minifields with a 90% accuracy using the July 16th and August DIFX and GLTX image data. Though Figure 112 does not indicate the fact, the AHTY weather data was considered to be an insignificant yield predictor and was not included in the yield prediction equation! The GLTX image data, due to its high correlation with weather data, was selected for inclusion into the yield prediction equation instead of weather data.

The DIFX and GLTX image data set demonstrated itself to be a superior yield predicting data set than that defined by the variables YSAX and GLTX. The DIFX/GLTX data set was able to account for 85% of the total yield data variance; whereas, the YSAX/GLTX data set could only account for 79% of the yield data variance at the 90% confidence level. This result is not surprising when considering that the multivariate analyses, pages 259-264 , identified the GLTX and DIFX variables as primary and secondary components respectively. Another reason why the

329

result is not too surprising is because the interpretation of

Regression #8 indicated that the DIFX data set accounted for

more yield data variance than the YSAX image data during the

July 16th through August period.

Interpretation of Regressions #9, #10, & #19 (Ref: Appendix G)

The reader may have noticed during the preceding discussion of the regression studies, that three regression studies, #9, #10, and #19, were excluded from consideration. An elaborate technical discussion of these studies was omitted for several possible reasons; for example, if the study was determined to be redundant with respect to another study, or if the extractable information was of little importance to warrant a detailed elaboration. This section briefly describes the objectives of each of the three rejected regression analysis studies.

The tenth regression study performed a forced-forward regression (FFR) of the GLTX image data in an attempt to predict the average yield of the entire site, YLDT. The results of the study were phenomenal in that 97% of the yield data variance could be accounted for as early as July 16th. However, the data set consisted of only the image data and yield data of the three sites (a far cry from the 11 to 16 minifields' data used in the other studies). Though F-table charts existed which, when interpreted, assured this investigator that the yield of the three sites was truly predicted as early as July 16th with a 90% accuracy, this investigator was still wary. If anything positive could be said for the experiment, the results of this

331

study did not contradict the interpretations based on the data of the 16 minifields. Perhaps in a future study, the feasibility of predicting the yield of an entire field, based upon the field's average albedo during the May thru July period, could be investigated. However, at this time, the author has chosen to make no conclusions from this study's results.

The ninth regression study was originally performed to investigate the feasibility of yield prediction as a function of the key variables which were associated with the principal components identified in the multivariate analyses. Three principal components had been identified in the multivariate analysis studies, pages 243-71, which could be identified with the variables GLTX, DIFX, and weather data respectively. The failure of the multivariate analysis study was that its results could not be used to interpret the ability of each principal component's ability to predict yield. Consequently, this particular regression study attempted to analyze each principal components' variables, to evaluate how well the particular data sets could account for the yield data variance (predict yield). However, the author found that similar data sets had been evaluated in other studies. Since the other studies' results seemed to clearly illustrate how good yield predictor

332

variables identified with the principal components determined
during the multivariate analysis study, the interpretation of
the ninth regression study was judged to be potentially
redundant. Regression #9 is mentioned at this point to simply
reassure the reader that this investigation of the data set
was not overlooked.

The final regression which had been omitted from any
previous discussion was the nineteenth regression analysis
study. Two data sets of image and weather variables from the
July 16th and August collection periods had been selected
using two criterias. The first data set was comprised of
variables which were identified as constituents of the best
primary and secondary principal components identified in the
multivariate analysis study. The second data set was com-
prised of variables which were identified as significant
yield predictors in preceding regression analysis studies
discussed previously. Obviously, such an experiment was guided
by curiosity rather than being an attempt to extract useful
information relevant to yield prediction. Consequently, the
detailed description of this regression study and the inter-
pretation of its noninterpretable results were omitted from
this report.

Conclusions of the Feasibility Study in Crop Yield Prediction

The conclusions of the regression analysis studies indicate that accurate and timely agricultural crop yield prediction is feasible through analysis of remotely-sensed data; specifically, a temporally-registered, multispectral image data set. That is, crop yield may be predicted through analysis of the crop's changing multispectral albedo during a dynamic portion of the growth cycle (i.e.: pollination thru post-pollination phase).

Several general conclusions were derived from the interpretation of the results of the multiple regression analysis studies. The studies were performed to investigate the feasibility of crop yield prediction utilizing only temporally-registered multispectral image data. The conclusions address several subjects that are pertinent to crop yield prediction; including: optimal time periods for multispectral imagery collections for yield prediction, optimal spectral bands, optimal types of image data, the significance of weather data, etc. Though specific conclusions were presented in the preceding interpretations of the regression analyses, those detailed observations were condensed to the following general conclusions:

1. The feasibility of timely (as early as pollination phase) and accurate prediction of agricultural crop yield (87% of the actual crop yield variance was accounted for with a 90%

334

accuracy), for fourteen small corn fields in Central Pennsylvania, was demonstrated utilizing a yield prediction equation that was solely dependent upon remotely-sensed data; specifically, a digital, temporally-registered, multispectral image data set.

2. When compared to the variables defining the time-variant, multispectral image data, the weather variables were observed to be insignificant predictors of agricultural crop yield.

3. Though multispectral image data sets, collected on 29 April, 2 July, 16 July, 19 August, 4 September, and 8 October 1974, were analyzed to determine the feasibility of crop yield prediction, the variables defining the imagery collected during mid-July and mid-August were prime predictors of agricultural crop yield. Optimal multispectral imaging periods for accurate yield prediction of the same Pennsylvania test sites may actually occupy some subset of the latter time interval (Regressions 1-6,7,8,13 and 15).

4. Though the two temporally-registered multispectral image data sets of mid-July and mid-August were found to be adequate for timely and accurate yield prediction, the analysis of three or more image data sets, spanning from early July to early September, did increase the percent of yield data variance accounted for to nearly 100% with a 90% accuracy (Regressions 15-18 versus 7,8).

5. The yield analysis from two or more bands of multi-spectral imagery, registered over time, provides better prediction results than when only one band's imagery is utilized. The optimal two-band yield predictor combination was the green/near-infrared image data combination. Highest accuracies in yield prediction occurred when image data from all three bands was analyzed (Regression 1-6 versus Regressions 7,8,11,12).

6. The average density of the digital multispectral images of the three large corn agricultural test sites, defined by the variable "GLTX", was highly correlated to weather data. In a regression equation for yield prediction, GLTX could be substituted for weather variables with no decrease in the prediction accuracies (Regression 18).

7. Some image variables were found to be better yield predictors than others. Surprisingly, the image variagle "DIFX", created to empirically remove the effects of variability in image density due to film processing and exposure so that monthly multispectral imagery data might be linked by some commonality feature, provided less yield prediction information than the variable defining the actual digital, non-adjusted grey level, "YSAX". This discrepancy could not be explained and should be resolved in a follow-on study (Regressions 1,3,5,7,11 and 13 versus Regressions 2,4,6,8,12 and 14). Interestingly enough, when the DIFX data was combined with GLTX

336

data in a regression equation for yield prediction, the results were excellent and superior to the results achieved with the YSAX variable in combination with the GLTX data. This result was in agreement with the multivariate analysis study which identified the GLTX variables and DIFX variables as the main constituents of the primary and secondary principal components respectively (Regression 18).

Semiautomatic Crop Identification Using Multispectral Imagery

Introduction

The preceding sections indicate that accurate yield prediction, early in the growth cycle, is feasible using temporally-registered, multispectral imagery. Yield prediction of fourteen minifields in Central Pennsylvania was demonstrated with a 90% accuracy using the multispectral imagery collected during the pollination stage of the corn growth cycle (July 16th thru August 19th). Despite these promising results, unless a crop can be identified by using only multispectral imagery, crop yield prediction models, which require the information of a temporally-registered multispectral image set, will not be feasible.

This final section discusses the experiments which were performed to investigate the feasibility of semiautomatically identifying an agricultural crop, corn, utilizing digitized and registered multispectral image data sets. The experimental program was initiated with the prior knowledge that the results would be marginal at best, because the narrowband filters of the multispectral camera had not been selected for speciating corn; but rather, to detect stress levels within corn during the entire growth cycle (Ref: Pages 50 - 53). Before proceeding with a

338

discussion of the semiautomatic crop identification study, a
literature review is presented to familiarize the reader with
different crop identification techniques that have been investi-
gated.

Literature Review: Agricultural Crop Identification

Introduction:

Accurate agricultural crop identification, crop yield
measurement, and yield prediction utilizing remotely-sensed
data are a few of the significant objectives of the national
remote sensing programs. If the feasibility of crop yield
prediction using remote sensing techniques is to be demonstrated,
the associated remote-sensing system must be capable of iden-
tifying the subject crop with a high degree of accuracy. Such
a system would consist of a sensor, a data storage medium, and
a processing capability (i.e., computer, image interpreter).
The processor would analyze the imagery and generate statistics
that would accurately measure the acreage of the subject crop
that was imaged in the "sensed" field of view.

Though many sensors exist today which have some potential
for detecting and identifying specific agricultural crops, the
identification accuracies have been generally unacceptable for
supporting the crop yield prediction requirement. These sensors
comprise three groups whose bandwidths range from the visible
microwave regions of the electromagnetic spectrum: photo-
graphic, electro-optical scanners, and radar. Radar sensors
have been only marginally investigated and the conclusions

340

concerning their ability to speciate crops are indeterminant at this time. However, both photographic and scanning sensors have demonstrated quite promising results in solving the crop identification problem. This literature review synthesizes the work and significant discoveries by researchers employing photographic and scanning remote sensing systems for crop identification ((Bauer,(5) provides a concise review of the remote crop identification technology)).

Crop Identification Utilizing Photography:

Prior to the development of multispectral electro-optical scanners during the 1960 decade, crop identification studies relied on data extracted from photography acquired from airborne camera systems. Aerial film types commonly utilized included black and white, black and white infrared, and both color and color infrared ektachrome. Cameras employed in the studies were typically single lens mapping cameras. However, near the end of the decade, the development of filtered, multiple lens cameras or "multispectral" cameras came into being (i.e., Itek Corporation - Nine lens camera; I^2S Corporation - Four lens camera; Spectral Data Corporation - Narrowband Four lens camera). The physical basis for the multispectral cameras was the fact that agricultural crops or any other material had a unique spectral signature (Yost and Wenderoth(54), Pavlin(41)).

341

This meant that, with some exceptions, if a material was analyzed with a spectroradiometer, its recorded reflectance would be distinct from other materials and uniquely defined as a function of wavelength (and time in the case of dynamic materials).

During the corn blight experiment of 1971, the ability of a photointerpreter to analyze film and identify agricultural crops with reasonably high accuracy, 75% to 93%, had been demonstrated (MacDonald,33). However, other studies demonstrated that the interpreter's accuracies depended upon the crop type, time of the growing season, and the scale of the imagery (Poulton,42; Manderscheid,35). After considering the untimeliness of the photointerpretation approach to remote crop identification and the mediocre to exceptional accuracies, remote-sensing scientists were forced to look at computer technology for a solution.

If film data was to be analyzed by a computer, the process of digitization was required to convert the film density data to computer recognizable numeric data. The product of the digitization process was a digital image stored in computer memory which was comprised of an array of numbers whose value might typically vary from zero (dark density value) to 255 (bright density value). The numeric array could then be displayed on television, processed, enhanced, or analyzed. After

342

the digitization of an imaged scene of an agricultural region, a computer could be trained to recognize a crop based on three general measures of features that could be extracted from a digital image. These measures are characteristics of the subject crop and include the digitized density or tonal value, texture, and any spatial features. A multispectral data set, consisting of "n" images, once digitized, would allegedly offer "n" times the number of useable feature statistics for automated crop identification.

The Problems of Crop Identification Using ERTS MSS Imagery:

When the National Aeronautics and Space Administration's earth resources satellite (ERTS-A) first began transmitting data, agricultural scientists were hoping that the additional spectral information gathered from the satellite's multispectral scanner (MSS) might provide the key for classifying agricultural crops on a national scale. Unfortunately, scientists discovered that the ERTS satellite could not provide the panacea for the problems of automated crop identification using remotely sensed imagery. The bandwidths of the ERTS-A MSS were limited to relatively broad bandwidths; Band 4 - Green (0.5-0.6u); Band 5 - Red (0.6-0.7u); Band 6 - Near infrared (0.7-0.8u); Band 7 - Near infrared (0.8-1.1u). Though this was a reasonable selection of bandwidths for a first earth resources satellite, the

343

selection was attempting to satisfy as many resource or environ-
mental problems as possible. Though many papers have been
written which review exploitation studies of the MSS imagery,
it seems that they are best suited for general tasks; i.e.,
mapping and measuring water bodies, differentiating hardwoods
from softwoods, etc. The reason the MSS exploitation has
fallen short of identifying agricultural crops reliably is due
to the fact that agricultural crops are spectrally distinct from
each other in only very small bandpass intervals of the visible
and near-infrared spectrums and these differences change
dynamically with the maturation of the crop (Yost & Grodewald,
20).

During the analysis of the ERTS-A MSS imagery, several
physiologically-related problems impeded the ability of the
computer to accurately identify crops. For example, Roberts(45)
and Earth Satellite Corporation(1) noted that crop speciation
was a periodic function of the growing season during which crops
were being identified. For example, during the early part of
the maturation cycle, when the leaf to soil ratio was found to
be less than twenty per cent, the plants could not be resolved
by the system and automated crop identification was not possible
(Roberts,45). D. H. Von Steen(53) noted that in the analysis of
only one ERTS-A frame, crop identification was not consistent

344

from one area of the scene to another due to differences in
weather, season, hybrid, and cultivation practices. He further
noted that to train on data in one area of the image frame to
identify a crop in another area of the frame was useless.
Safir(38) also found crop identification to be a function of the
growth stage of the crop. He was best able to differentiate
forests, corn, and soybeans when the respective crops had their
densest green canopy. Safir had difficulties with the semi-
automatic identification of senescent vegetation and noted that
when performing digital crop identification, that edge-effect
problems were minimal only when the atmosphere was clear and
fields were large.

Besides the physiological reasons for the suboptimal crop
identification accuracies using advanced image processing tech-
niques, some scientists found fault with the ERTS MSS system and
the ground truth programs. Earth Satellite Corporation(1)
observed that part of the reason for the inaccuracies in using
the ERTS MSS data for crop identification was the lack of
adequate ground truth to support the analysis. Accuracies were
often a function of an image interpreter's decision or supporting
image truth. Haralik(8) identified other factors negatively
affecting crop identification by computer. Since the ERTS
imagery was often delivered to investigators via a film medium
(produced from ERTS MSS analog or digital tapes) which required

345

digitization for computer analysis, the final automated crop identification accuracies became a function of the lack of calibration of the film, film processing, and the age of the film. Roberts(45) commented that simply increasing the number of bands on the ERTS satellite sensor would not nullify the crop identification problems because in past studies it had been determined that there was only a 2% to 3% increase in identification accuracies using twelve bands as opposed to four or five bands.

Specific Studies in Crop Identification Using ERTS MSS Imagery:

Generally, the many studies performed for semiautomatically identifying crops using ERTS imagery differ in the geographic area to be classified, the types of crops to be identified, and the computer logic used in the classification. For example, Haralick(8) has developed texture feature discriminants and Bayesian decision rules for his studies in crop identification with ERTS imagery. Von Steen(52) implemented the Pennsylvania State University classifier, which is basically a linear discriminant classifier, in his study of automated crop identification. Furthermore, computer classification algorithms and decision logic are being developed at many universities, corporations, and government installations which ultimately could be used for automated crop identification with ERTS or other varieties of imagery. Some of the more elaborate systems

346

dedicated to imagery processing include the LARS System of
Purdue University, the KANDIDATS System of the University of
Kansas, and the DICIFER System of the Rome Air Development
Center (RADC) Image Processing Facility.

There have been many studies of both ERTS MSS imagery and
photographic imagery collected at both low and high altitudes.
The following are just a few which made significant contri-
butions in the field of remote crop identification: Minter(37),
Bauer(5, Review), Poulton(42), Safir(38), Thompson(51), Draeger
(14), Bauer and Cipra(4), Baumgardner(6), Davis(12), Downs(13)
and (1). Unfortunately, the literature also contains infor-
mation that could confuse researchers as to how well crop iden-
tification is being performed. For example, some authors
attempt to base their accuracies on classes they identify to be
crops but which are not recognized universally as typical
agricultural crops; i.e., water, stubble, burned crops, turned
crops (Bizzell(7)). Or the classification may be over an area
which consists of only a few major crops such that the chance
of the computer logic misclassifying is remote (Horton,24).
Other authors might publish their high classification rate after
analyzing only one ERTS image which happened to be acquired
during the optimal discriminating period for the subject crop
(Johnson,28; Horton,24; Safir,38; and Draeger,14). It should be
pointed out that no known paper has concluded that accurate

347

automated crop identification is possible throughout the growth
cycle for all major crops using ERTS MSS or any other satellite
film or electro-optically generated imagery.

The Uncanny Success of the Narrowband Multispectral Sensor for Crop Identification:

There is only one known study that identified the capability
of a sensor system to identify an agricultural crop throughout
the growth cycle without being a function of geography, altitude,
or the dynamic nature of living vegetation's spectral signature.
A narrowband multispectral camera system complemented with
additive color display technology was able to speciate opium
poppies from all background materials at low and high altitudes,
over the entire growth cycle, in three different areas of the
world (Haynes,23). This sensor system's success is attributed
to the use of specific narrowband filters selected from the
analysis of visible to near-infrared radiometry measurements of
opium poppies and background materials which had been collected
over the crop's entire growth cycle. The filters of the system
were selected in the regions of the visible and near-infrared
spectrums where the differences in spectral reflectivity between
the subject crop and background or surrounding vegetation is
maximized, Wenderoth(54). The results of the photo-interpreta-
tion analysis of the additive color product demonstrated a 98%
classification accuracy of the subject material.

348

Exploitation of the Temporal Fluctuations of Agricultural Crops'
Dynamic Spectral Signatures:

Other scientists have observed crop identification to be
a function of the phase of the crop growth cycle at the time of
overflight. These scientists have evaluated ERTS imagery with
both photointerpreters and computers to determine if there was
significant improvements in identification accuracies when the
analyzed image set was collected at specific times of the growth
cycle. This premise was studied and verified by MacDonald(33),
Poulton(42), Safir(38), Baumgardener(6), Mahlstede(34), Downs
(13), Steiner(49), and Roberts(45). From these initial studies,
other investigators established the value of combining both the
spectral and temporal dimensions of ERTS MSS imagery for
increasing automated crop identification accuracies.

To add the temporal dimension to a multispectral imagery
data set requires a registration process which produces a series
of multispectral images whose picture elements (pixels) would
correspond to the exact same location on the ground. ERTS MSS
data is a good example of registered imagery. For a given frame
of ERTS imagery, each picture element associated to a given
spectral image corresponds exactly to the ground imaged by the
same picture element of another spectral image. Spectral images
within an ERTS MSS frame are already registered due to the
nature of the sensor. However, to register an ERTS frame imaged

349

in May with an ERTS frame imaged over the same geographic area
in October is quite difficult unless the orientation parameters
of the ERTS satellite platform were the same on the two passes.
Fortunately, computer software algorithms exist which perform
the registration task. However, the registration is seldom
exact and the actual pixel registration may be in error by
one or more pixel positions. Nevertheless, Carlson(9),
MacDonald(33), Bauer(4), Baumgardner(6), (1), Mahlstede(34),
Davis(12), Steiner(49), and Roberts(45), have either suggested
or actually performed a temporal registration of multispectral
images and discovered that the accuracies of crop identification
increase significantly. In one example, Bauer(4) utilized the
latter approach to speciate corn and soybeans over a two
thousand square mile region with an 83% accuracy using ERTS MSS
imagery and automated image processing techniques.

Generally, crop identification studies have reported
accuracies which varied from submediocre, when considering
multiple crop classifications, to spectacular when considering
only one crop. For example, Davis(12) compared the image
processing systems of the Environmental Research Institute of
Michigan (ERIM) with the LARS System of Purdue University.
Average classification accuracies of ERTS MSS nontemporally
registered imagery was 67% for corn, 64% for soybeans, and 60%

350

for other crops (average classification accuracies of the two image processing systems combined as a function of crop type). Interestingly enough, the average for these two image processing systems generated higher classification accuracies when analyzing a temporally-registered ERTS multispectral data set; i.e., 82% for corn, 74% for soybeans; and 59% for other crops. Back in 1973, when just single frames of multispectral imagery were being analyzed, low overall accuracies were not too surprising; i.e., 34% overall crop identification accuracy, Steiner(49). For nondigital photointerpretation of Skylab and ERTS imagery, overall crop identification accuracies of 75% were quoted by Poulton(42). In some studies, some crops were significantly more recognizable on single frames of ERTS MSS imagery; i.e., rice, with a 91% accuracy (Poulton,42), and wheat in Montana with an associated 90% accuracy (Flores,16).

Unfortunately, the reported accuracies were simply not good enough to satisfy the crop identification requirements of crop yield prediction programs using remotely-acquired imagery. The investigations did establish the potential of multispectral, temporally-registered imagery for satisfying the requirement. If a system was to be developed for crop identification and yield prediction, it seemed that the sensor should be multispectral and provide the maximum amount of information about

351

the subject crop in terms of imaged tone, texture, and spatial features. Certainly, the MSS system of the ERTS satellite was thought to meet these requirements except in one important respect; the detection bands of the ERTS MSS were not dedicated to speciating one or more particular crops.

Literature Review Conjecture:

As mentioned previously, the only system that had any success at identifying a crop with near 100% accuracy over the crop's entire growth cycle was the narrowband multispectral system developed by Spectral Data Corporation of Hauppauge, New York (Haynes,23). In that study, radiometry of opium poppies and typical background vegetation had been performed over the entire growth cycle of poppies in three different locations of the world. The resultant analysis of the spectral signatures from 400 Nm to 900 Nm revealed intervals of the spectrum where the crop could be identified from the background material. Narrowband filters selected in the predetermined spectral intervals were the key to the system's success. The crop identification feat could be performed at low and high altitudes at three different locations of the world over the entire sensor-resolvable stages of the poppy growth cycle (e.g., if the sensor could not resolve the vegetation, it could not detect it). Though this success may not be repeatable (?) for such crops as corn, soybeans, and wheat, the technique should be

352

investigated. Perhaps the narrow-band multispectral imagery combined with electro-optical technology and temporal registration techniques provides the true key to semiautomated crop identification for crop yield prediction.

Semiautomatic Crop Identification: Solutions to the Problems

The feasibility of accurately predicting the yield of a crop early in the growth cycle has been demonstrated through regression analysis of a temporally-registered multispectral image data set. If these significant results are verified in future studies, future crop yield prediction models, requiring temporally-registered multispectral imagery of a crop during two or more intervals of the growth cycle, will also require that the crop be semiautomatically identified using the imagery. Semiautomatic crop identification is a necessity of any model which requires temporally-registered multispectral images as the principal data set for yield prediction.

The preceding literature review of crop identification studies indicates that only two techniques are known to demonstrate a good to excellent ability to speciate crops. The most successful crop identification technique which demonstrated excellent accuracies, utilized a multispectral camera system having narrowband filters that were selected to speciate a crop from background materials. The second technique which successfully identified crops utilized the temporal parameter of temporally-registered multispectral image data sets and sophisticated image processing techniques.

Perhaps it might seem that semiautomatic crop identification is controlled by the sensor system and the image

354

processing software capability. However, sophisticated image processing software is not the critical requirement for accurately identifying a crop semiautomatically. Semiautomatic crop identification is primarily a function of the features which can be extracted from the digital imagery (i.e., tone, texture, etc.), that can be used to semiautomatically separate the classes (i.e., corn, soybeans, wheat, water, etc). Regardless of the sophistication of the pattern recognition software, if corn "looks" like wheat or oats in the multispectral imagery, the computer will be unable to speciate the crops. Similarly, if the crops' image features appear identical on two successive but temporally-independent multispectral images, the computer will again be unable to separate the classes. Simply expressed, if the sensor cannot "see" (detect) the crop, then the computer will be unable to identify the crop. Therefore, considering the importance that the sensor plays in the semiautomatic crop identification process, every attempt should be made to design the remote sensing system to detect the crop (i.e., narrowband multispectral camera system technique).

The exploitation of the temporal dimension of multispectral image data has also been demonstrated to be extremely useful for increasing the semiautomatic crop identification and crop yield prediction accuracies. Capitalizing on this factor's value can

355

only improve the ability of the sensor-image processing system to semiautomatically detect the subject crop.

There may be an optimal mix between the technical requirements of future crop identification and crop yield prediction systems. Since it is well known that crops demonstrate a dynamic spectral reflectance that changes during the course of the growth cycle, there is no evidence that the narrowband multispectral technology will work equally well on all economically important crops throughout the growth cycle. However, if certain time periods were identified for collecting multispectral imagery for predicting the yield of the crops, then the narrowband multispectral sensor's bandpasses could be selected to speciate the particular crop during the particular period. In the latter case, the advantages of the narrowband multispectral sensor and the temporal registration of the imagery would be utilized to insure that the crop was identified with excellent accuracy. High crop identification accuracies, completed by the accuracy of an imagery-dependent yield prediction model, would result in optimal yield predictions of the subject crop.

The success of the narrowband multispectral camera in identifying an agricultural crop, opium poppies, under adverse conditions (i.e., high or low altitude collection, various stages of the growth cycle, and various locations of the world)

356

was due to the fact that the bandpasses of the multispectral filters were selected to occupy regions of the visible and near-infrared spectrum in which the subject crop was most distinguishable from the other background crops. If the filters of the camera system, or the bandpasses of a future multispectral scanner system, were selected to detect stresses in crops, instead of speciating the crops, the multispectral imagery would demonstrate a marginal capability as a prime data set utilized to semiautomatically identify the crop. Though the latter statement may be logical, no proof of its validity has yet been demonstrated. Consequently, in order to indirectly substantiate the importance of selecting correct bandpasses for a narrowband multispectral sensor dedicated to semiautomatic crop identification and yield prediction programs, a series of experiments in semiautomatic crop identification were performed using the stress-detecting multispectral imagery collected during the course of this study. Specifically, corn speciation was attempted using additive color display techniques, semi-automatic image processing techniques, and the multispectral imagery dedicated to stress detection. If conscientious image processing of this imagery results in suboptimal semiautomatic crop identification accuracies, the theory that the narrowband multispectral bandpasses of the sensor must be dedicated to crop speciation rather than stress detection will be sub-stantiated by an indirect verification.

357

Speciation of Corn Using Additive Color Analysis Techniques

Multispectral photography of July 16th was analyzed with additive color display techniques (Ref: pages 54 - 63) to speciate one particular species of vegetation, corn, from other types of vegetation. Observing the dynamic nature of the spectral reflectance of corn and other crops through analysis of the crops' spectral reflectance signatures (Reference: Appendix A, page 391), it was determined that the July 16th multispectral imagery collection period had the best potential to demonstrate the speciation of corn from other crops. Due to the varieties of vegetation present, multispectral imagery of a section of the Pennsylvania State University Agronomy Farm was chosen to evaluate the sensor's potential to speciate corn from other agricultural crops.

Figures 113 and 114 illustrate the fields of interest. The field annotated as "a" in Figure 113 is completely covered by a dense growth of "normal" healthy corn. When this field was compared to field "a", it was observed that the fields' colors were the same except that a pattern existed in the "a" field. This pattern was attributed to a high percentage of low level weed cover (verified from the ground truth survey). However, field "a" was still identifiable as corn because of its characteristic color in the multispectral composite image, Figure 114.
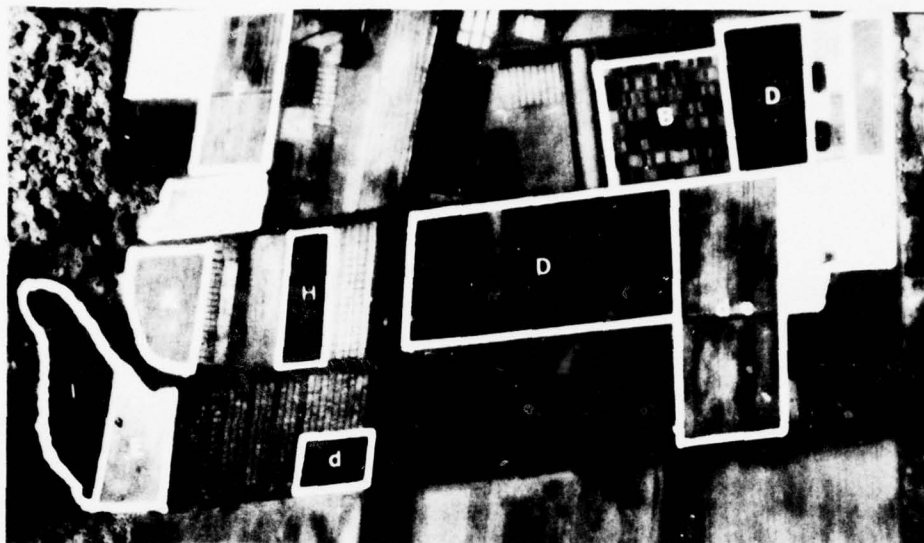
Figure 113 Annotated image truth photograph of the same area as shown in Figure 114.

FIELDS	CROP AND DESCRIPTION
A	This is a normal corn field.
a	This is a "No Till" corn field, it has a high percentage of weed cover at ground level.
B	This is a field containing ordered replications of corn, oats, crown vetch, soybeans, and wheat. The individual plots are approximately 10' x 20'.
D	Oat stubble, the oats in this field were harvested before the photograph was taken.
E	Alfalfa
de	This field contains a combined planting of oats and alfalfa, it must be considered an experimental field.
d	Headed oats.
F	Wheat
G	Potatoes
H	This field contains a combined planting of wheat and timothy, it must be considered an experimental field.
I	Pasture grass and timothy.

Note: Some of the other experimental fields do contain corn, however, the plots are small and not in the normal stage of growth for this time of year.

359

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

visual presentation of the ... field ...
with the multispectral rendition, Figure ... it can be seen
that no other crop at the Agronomy Farm other than corn
displays exactly the same color. When examined on the additive
color viewer screen, the non-corn field that was most similar
to corn field "a" in color was that indicated by the field
labelled "b" in Figure 113. The "b" field was an experimental
field which combined the planting of alfalfa and oats.

Figure 114: Photographic reproduction of an additive color
viewer screen presentation of part of the Pennsylvania State Uni-
versity Agronomy Farm, illustrating the speciation of corn from
other crops. Corn fields and test plots exhibit a characteristic
brown-red color. The ground truth for this area is shown in
Figure 113. Mission GR74-17C, frame 41, 15 July 1974, Altitude
10,000 feet AGL, Rendition: The 541-591 nm (green) band is
imaged as green, the 620-680 nm (red) band as blue, and the
810-885 nm (near-infrared) band as red.

Though the stress-detecting multispectral imagery demon-
strated an apparent ability to speciate corn from other crops
using additive color analysis techniques, the imagery and
technique failed to speciate corn during most of the growth
cycle. Figures B-1, B-3, & B-5 of pages 411 - 416 respectively
illustrate the ability of the multispectral filters to detect
levels of stress within corn. Healthy corn typically appears
as a light purple and varying degrees of stress appear as gold,
blue, tan, etc. The filters inability to speciate corn is
illustrated in most color renditions whereby corn, trees, and
pasture grass appear to have the same hue (i.e., Figure B-7,
page 418).

Despite the (expected) poor results of the filters to
speciate corn from other crops, an investigation was performed
to study the feasibility of semiautomatically speciating corn
using registered multispectral images and the RADC DICIFER
Image Processing System (Ref: page 117). This study had the
objective of demonstrating the inability of advanced image
processing software to speciate a crop using narrowband multi-
spectral image data which was not dedicated to speciating the
subject crop.

361

Digital Image Processing Procedures for Crop Identification

The information of a photographic image consists of its tone, texture, and spatial features. Since materials reflect light uniquely and the reflected light is recorded as a density on film (or a voltage signal on an electro-optical scanner's storage medium), the tone features of the imaged crop have the potential to be the principal measurements utilized to speciate the crop. For example, if the multispectral imagery of this study had been collected with filters selected for speciating corn, the resultant imagery would be expected to have a unique tone quality in each spectral image when corn was imaged by the multispectral camera.

Since any density of a photograph can be converted to a numeric density value and stored in magnetic memory using a computer digitizer, materials imaged on digitized film may be characterized by the statistics or numeric density values that define the material class (i.e., corn, water). Naturally, registered multispectral images provide "n" times the number of features for discriminating the corn from other crops. The DICIFER Image Processing System (Ref: pages 117-145) can utilize these tone features (and any texture features) that are extractable from the digital multispectral images to semiautomatically identify the corn from other crops.

This subsidiary feasibility study in semiautomatic crop identification utilized the same techniques to digitize, register, and classify a multispectral scene as those which were used in the study to semiautomatically map stresses within corn fields (Ref: page 200). Figure 66, page 204 , summarizes the interactive procedures which were followed during the semiautomatic crop identification study.* Basically, the interactive process consisted of displaying the multispectral images; creating a design set which sampled the features of the imaged classes (i.e., corn, trees, water, etc); evaluating the image features' (i.e., multispectral tone, texture) ability to differentiate a class; selecting a classifier logic (i.e., Fisher, Boolean, One- or Two-Space logic); classifying the entire scene automatically using the multispectral features and logic which were judged to be acceptable for differentiating the classes of the design set; checking the results and modifying the procedures if necessary.

----

*Note: For the sake of brevity, the image processing procedures which were explained in detail in previous sections, which described the interactive process for semiautomatically classifying a digital image set, are not described in this section of the report. The reader is presumed to understand the particulars of digitization, registration, creating a design set, etc. Appropriate references to sections which elaborate on concepts, hardware, techniques, etc., are included for the reader's convenience. A basic understanding of the image processing steps can be extracted by reviewing pages 200--214. These latter techniques were used to semiautomatically map stress levels in imaged corn fields for yield measurement purposes.

Though two multispectral image data sets were considered in the semiautomatic crop identification investigation, the analyses of only one data set are presented in this report. A multispectral photograph of Sites Six and Seven, collected on 22 July 1974 (Mission: GR74-52C,frame 6) at an altitude of 50,000 feet AGL, comprised the primary data set. The second image data set consisted of a high altitude multi-spectral photograph of Site One of the Pennsylvania State University Agronomy Farm. Figures 115 and 116 illustrate the green and red digitized images of the primary data set respectively. Figure 117 presents the near-infrared image of the primary data set which was digitized and registered to the green and red images depicted in the preceding figures.

Various design sets were created which sampled several classes of corn, trees, and other general classes of interest. Figure 117 illustrates the locations of the various area files of the design sets which sampled image features of the classes which were to be differentiated (Reference Figure 118 for ground truth of the imaged scene). These design sets' tone and texture features were evaluated and the selected features were utilized to classify the entire scene or "test set". The resultant "decision images" represent the final product of the computer. These images effectively provide a digital map of all of the

364

Figure 115: Green Image of Sites 6 & 7 which was digitized and registered with the red and near-infrared images from the same frame of multispectral photography. Photography was originally collected at an altitude of 50,000 ft (AGL) with an AFFTC U-2 aircraft on 22 July 1976 (Mission: GR74-52C, Frame 6). This image was utilized in the spectrally-registered image data set for investigating semiautomatic crop identification.

Figure 116: Red Image of Sites 6 & 7 which was digitized and registered with the green and near-infrared images from the same frame of multispectral photography. Photography was originally collected at an altitude of 50,000 ft (AGL) with an AFFTC U-2 aircraft on 22 July 1976 (Mission: GR74-52C, Frame 6). This image was utilized in the spectrally-registered image data set for investigating semiautomatic crop identification.
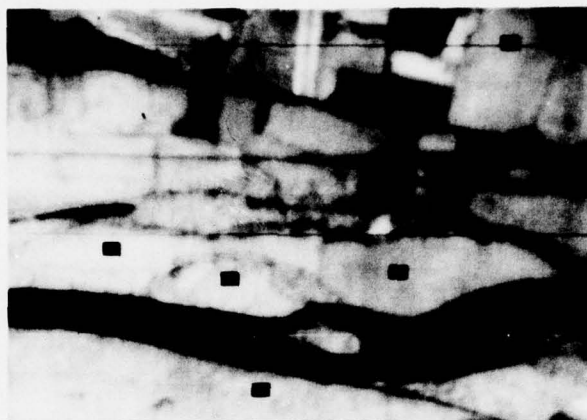
Figure 117: Near-infrared digital image of Sites 6 & 7 which was registered with the green and red images from the same frame of multispectral photography. Photography was originally collected at an altitude of 50,000 ft (AGL) with an AFFTC U-2 aircraft on 22 July 1976 (Mission: GR74-52C, Frame 6). The small black boxes mark the locations of the area files which sampled tone and texture features of the imaged classes of corn, trees, water for the purposes of semiautomatically identifying corn in the entire imaged scene. Reference Figure 118 for the ground truth of the scene.
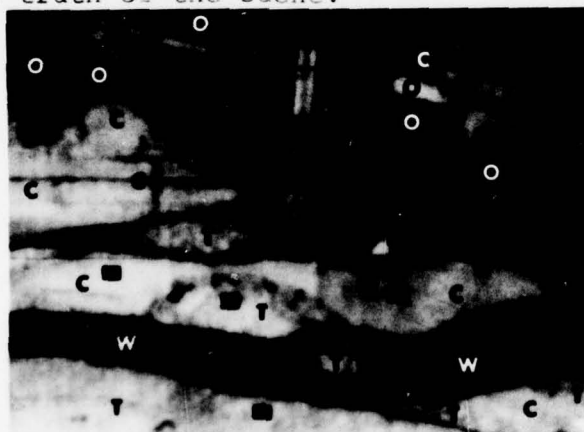
Figure 118: High altitude digital image of Sites 6 & 7 (IR spectral image) with annotated classification of the imaged scene (ground truth): C-corn; T-trees; W-water; and O-other. In determining how well each decision image of the following pages identified corn, reference should be made to this figure.

366

identified corn, trees, water, and other classes imaged in the scene. The logic classifies each pixel of the test set based on correlation of the pixel's features with those of the classes defined in the design set.

The results of these experiments in semiautomatic crop identification are presented on the following pages.* The information of each decision image (i.e., Figure 119) includes a description of the features utilized in the design set; the classes being identified; the logic employed, and any important comments that would apply to the experiment. Furthermore, a brief comment is presented which qualitatively evaluates the success of the particular experiment to classify and identify corn from the other classes.

---

*Note: Each of the decision images was produced with the DICIFER System's Color Printer Device (Ref: pages 129   ). There may be some difficulty in discriminating two shades of yellow which corresponded to two area files of the design set which sampled trees as a class. Third generation photographic reproduction generally causes the differentiability of hue to be a difficult, if not impossible task.

Figure 119: Image Name-DDDD; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red; Trees-yellow; Water-blue; and
Rejects-purple. Reference Figure 118 for comparison.

Design and Test Set Spectral Images: G,R,IR of 22 July 74
Number of area files/class: Corn(3), Trees(2), and Water(1)
Area file dimensions: 81 pixels/area file
Number of features utilized: 3
Features: Tone
Classification logic: Fisher
Comments: This approach to corn speciation is probably the
most logical of all the investigations. Generally, if the
image data has significant information content for crop
speciation, excellent classifications should result.

Evaluation: This decision image was qualitatively judged to
demonstrate a fair classification of corn. Unfortunately,
though most of the corn was identified, too many trees were also
identified as corn (i.e., the trees bounding Site 6).

Figure 120: Image Name-CCCC; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red; Trees-yellow; Water-blue; and
Rejects-purple. Reference Figure 118 for comparison.

Design and Test Set Spectral Images: G,R,IR of 22 July 74
Number of area files/class: Corn(3); Trees(2); and Water(1)
Area file dimensions: 81 pixels/area file
Number of features utilized: 4
Features: Average Tone, defined by the equation:

$$M_1 = (M_1+M_2+M_3)/3,$$ where, for each pixel coordinate,
M_1 thru M_3 defined the density of
the pixel per each spectral image
of the design and test set;

**Texture features: Mean, Standard Deviation, and Range
measures extracted with a scanning 3x3 pixel box from
the design set area files of the** tone-transformed
image data **(Reference: Description of texture features
and the techniques used for feature extraction, pages
202-212).**

Classification logic: Fisher logic.
Comments: The new tone feature had the intent of representing
an additive density measure of the three spectral images,
to mimic the results of the analogue additive color results.

Evaluation: Relative to the other decision images produced
during the course of this study, this approach produced a
decision image that demonstrated a fair to good discrimination
of trees.

369

Figure 121:   Image Name-GOBN; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red; Trees-yellow; Water-blue; and
Rejects-purple.   Reference Figure 118 for comparison.

Design and Test Set Spectral Images: R,IR of 22 July 74
Number of area files/class: Corn(3); Trees(2); and Water(1)
Area file dimensions: 81 pixels/area file
Number of features utilized: 5
Features: Tone of the near-infrared and red images;
            Texture features: Mean and High density features
            extracted with a scanning 3x3 pixel box from the
            design set area files of the red spectral image;
            Texture feature: low density feature extracted
            with a scanning 3x3 pixel box from the design set
            area files of the near-infrared spectral image.

Classification logic: Fisher

Comments: The texture features were selected after 21 features,
which were extracted from the green, red, and near-infrared
spectral images, were evaluated using the "Probability of
Confusion" and "Discriminant" feature evaluation software of
the DICIFER Image Processing System (Reference: page 135).

Evaluation: This decision image demonstrated a relatively good
classification of corn in the peripheral fields surrounding
Sites 6&7. However, corn was confused with trees in Site 7.
The blocky texture is due to the nature of the texture feature
extraction algorithm which samples the texture via a scanning
3x3 pixel box (size is adjustable). Since appropriate areas
were rejected and corn was identified relatively well, this
image was considered to be one of the best resulting from the
series of crop identification experiments.

Figure 122:   Image Name-QQQQ; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red; Trees-yellow; Water-blue; and
Rejects-purple.   Reference Figure 118 for comparison.

Design and Test Set Spectral Images: R, IR of 22 July 74
Number of area files/class: Corn(3); Trees(2); and Water(1)
Area file dimensions: 81 pixels/area file
Number of features utilized: 2
Features: Texture features: High density (red image) and
           Mean density (near-infrared image) features extracted
           with a scanning 3x3 pixel box from the design set
           area files of the two spectral images.

Classification logic: Fisher*.

Comments: The two texture features were identified as being the
most significant of the 21 tone and texture features extracted
from the green, red, and near-infrared spectral images for dis-
criminating corn from trees (Ref: Comments of Figure 121).

Evaluation: This decision image demonstrated a relatively
good classification of corn (Estimated Accuracy: +80%).
Unfortunately, not enough trees are being identified and the
rejects are being classified as water. The texture features
utilized for corn identification are elaborate. Under normal
conditions, in which the multispectral imagery was dedicated
to crop speciation and not stress detection, a computer
dedicated to semiautomatic crop identification should be able
to discriminate corn from trees using only the tone features.

---

*Note: An attempt to produce a decision image using Boolean
logic and the same texture features resulted in a decision image
that was similar to QQQQ. The only difference between the two
images was that rejected regions were not identified as water,
but as rejects.

371
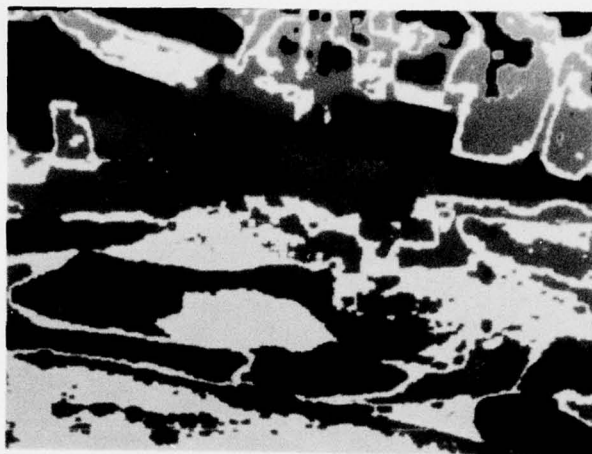
Figure 123: Image Name-RATS; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red; Trees-yellow; Water-blue; and
Rejects-purple. Reference Figure 118 for comparison.

Design and Test Set Spectral Images: R,IR of 22 July 74
Number of area files/class: Corn(3); Trees(2); and Water(1)
Area file dimensions: 81 pixels/area file
Number of features utilized: 2
Features: Texture features were identical to those used in
           Figure 122.
Classification logic: One-Space logic (Ref: page 136-38).

Comments: Reference: Comments of Figure 122.

Evaluation: Though this decision image demonstrates a relatively
good classification of trees and rejects, there are too many
misclassifications that make this decision image useless as a
crop speciation map (i.e., the water in the stream appears to
be classified in parts as both trees and corn; Site Seven has
been classified as a lake).

372

Figure 124:   Image Name-YYYY; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red; Trees-yellow; Water-blue; and
Rejects-purple.   Reference Figure 118 for comparison.

Design and Test Set Spectral Images: IR of 22 July 1974
Number of area files/class: Corn(3); Trees(2); and Water(1)
Area file dimensions: 81 pixels/area file
Number of features utilized: 4
Features: Texture features extracted with a scanning 3x3
            pixel box from the design set area files; features
            included: Mean, Median, Low density, and High
            density (Reference: Description of the texture
            features and the technique used to extract them,
            page 202-212).
Classification logic: Boolean logic; whereby, water was
            extracted first, followed by trees and corn.

Comments: The new features of texture were selected after 21
features, which were extracted from the green, red, and near-
infrared images, were evaluated using the "Single Space Histo-
gram" feature evaluation software (Ref: page 133-36).   The
features identified as best for corn discrimination were dif-
ferent than those determined "best" in the feature evaluations
using the "Probability of Confusion" and "Discriminant" feature
evaluation software.   Reference Figure 121 thru 123.

Evaluation: This decision image demonstrated only marginal
corn identification capability.   The decision image illustrates
some edge effect misclassifications (i.e., river).   (Ref: Figure
125).

373

Figure 125: Image Name-ZZZZ; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red; Trees-yellow; Water-blue; and
Rejects-purple. Reference Figure 118 for comparison.

Comments: This decision image was generated with the same
logic, features, and design set of Figure 124. The only ex-
ception to the similarity was that the Boolean logic identi-
fied the classes in a different order; that is, corn was
classified first in the test set, followed by the classes of
water, trees and rejects. This decision, image demonstrates
more edge effect misclassifications (i.e., "rejects" class
assigned to pixels bounding trees and corn). Many trees are
being confused with corn and vice versa (Ref: Site 7).

Figure 126: Image Name-ZORO; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red; Trees-yellow; Water-blue; and
Rejects-purple. Reference Figure 118 for comparison.

Comments: This decision image was generated with the same
features and design set of Figure 124. The approach to classi-
fication was different in that Fisher logic was utilized.
Though this decision image demonstrated promising results by
identifying corn in Site 7, it failed to identify corn in about
50% of the known corn fields bounding Sites 6 & 7.

374

Figure 127: Image Name-HURD; Corn Speciation Decision Image of Sites 6 & 7; Corn-red to orange; Trees-yellow; Water-blue; and Rejects-purple. Reference Figure 118 for comparison.

Design and Test Set Spectral Images: G,R,IR of 22 July 74
Number of area files/class: Corn(3); Trees(2); and Water(1)
   Note: Each sample of corn and trees was considered
   to be a separate class (i.e. past experiments
   have included the 3 samples of corn into the same
   class in the design set.
Area file dimensions: 81 pixels/area file
Number of features utilized: 7
Features: Tone features extracted from the G, R, and IR images;
   Texture features extracted with a scanning 3x3
   pixel box from the design set area files. These
   features were identical to those used in Figure 124.

Classification logic: Step 1 - Boolean logic separated two
   classes of trees from the other classes; Step 2 -
   Fisher logic was used to complete the discrimination
   of all classes.
Comments: Instead of incorporating "n" number of area files
of a given class into the design set of a given class, each
area file's sample was considered to be a separate class.
This approach was useful in determining which sample of corn
was being confused with trees, etc.

Evaluation: Comparing this figure to Figure 118, a deter-
mination of which categories of corn are being confused with
trees can be established: For example, the corn sampled in
Site 7 appears to be confused with the trees bordering Site 6.
It may be difficult to discriminate the two classes of trees
being identified because the two hues of yellow that identify
these classes are not distinct. The overall classification
of this scene was qualitatively judged to be poor.
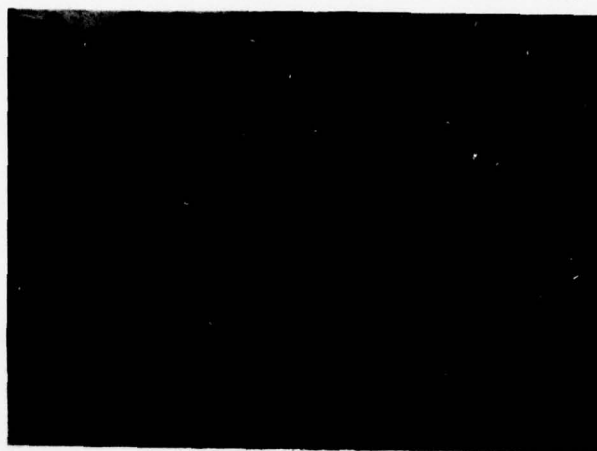
Figure 128: Image Name-SEAT; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red to orange; Trees-yellow; Water-blue;
and Rejects-purple. Reference Figure 118 for comparison.

Design and Test Set Spectral Images: G,R,IR of 22 July 74
Number of area files/class: Corn(3); Trees(2); and Water(1)
        Note: Each sample of corn and trees was considered
                to be a separate class.
Area file dimensions: 81 pixels/area file
Number of features utilized: 1
Features: Average Tone, defined by the equation:

$$M_1 = (M_1 + M_2 + M_3)/3, \text{ where, for each pixel coordinate,}$$
$$M_1 \text{ thru } M_3 \text{ defined the density of}$$
the pixel per each spectral image of the design and
test set.

Classification logic: Fisher logic

Comments: Reference Figure 127
Evaluation: This decision image was qualitatively judged to
demonstrate relatively poor classification of corn compared to
other decision images. Too much corn is being confused with
trees and vice versa. The intent of this figure was to simply
illustrate that regardless of the complexity of the features
utilized, if the crop discrimination information was not inherent
to the imagery, then the classes could not be differentiated
through any tricks of feature extraction or image processing.

Figure 129: Image Name-WORD; Corn Speciation Decision Image
of Sites 6 & 7; Corn-red to orange; Trees-yellow; Water-blue;
and Rejects-purple. Reference Figure 118 for comparison.

Design and Test Set Spectral Images: G,R,IR of 22 July 74
Number of area files/class: Corn(3); Trees(2); and Water(1)
        Note: Each sample of corn and trees was considered
               to be a separate class
Area file dimensions: 81 pixels/area file
Number of features utilized: 1
Features: Residual Average Tone, defined by the equation:

$$M_1 = (M_3 - M_1 - M_2)/3,$$ where, for each pixel coordi-
nate, M_1 thru M_3 defined the
density of the pixel per each spectral image of
the design and test set. Since the IR image was
always brighter than the other two images, its
greater numeric density M_3 was used to subtract
the density of the other two images. No scientific
theory was utilized to develop this measure. How-
ever, it may correspond to the additive density
measure of the three spectral images.

Classification logic: Fisher
Comments: Reference Figure 127
Evaluation: This decision image was qualitatively judged to
demonstrate a relatively good classification of corn in the
peripheral fields surrounding Sites 6&7; however, corn was
confused with trees in Site 7 and various regions of Site 6.
The intent of this figure was to simply illustrate that
regardless of the complexity of the features utilized, if
crop discrimination information was not inherent to the imagery,
then the classes could not be separated using any tricks of
feature extraction or image processing.

377

Semiautomatic Crop Identification Using Multispectral Imagery

Conclusions

The most significant result of the preceding image pro-
cessing experiments, was that semiautomatic crop identification
was demonstrated with only an estimated 50$\pm$30% accuracy, using
narrowband multispectral image data that was dedicated to
detecting levels of stress within corn. This rather poor
result was predictable and can be explained. First, consider
the fact that similar multispectral imagery, that had been
collected with the dedicated multispectral camera system,
provided an excellent data set for semiautomatic mapping of
crop stress levels and yield measurement (pages: 200-242   ).
The latter success was not surprising because the bandpasses
of the multispectral camera system had been selected to opti-
mally detect levels of stress within corn. The resultant
multispectral imagery's tone and texture features were optimal
for demonstrating semiautomatic stress level differentiation
within an imaged field of corn. Ironically, the same multi-
spectral camera system had demonstrated an excellent ability to
speciate crops when it had been previously utilized to identify
opium poppies throughout the resolvable periods of the growth
cycle, in various locations of the world, and over a range of

378

sensor platform altitudes. Consequently, it is concluded that the mediocre crop (corn) identification accuracies achieved with semiautomatic image processing techniques, was due to the fact that the narrowband multispectral camera system was dedicated to corn stress level detection and not to the speciation of corn.

The objective of this study was satisfied in that the experiment provided an indirect verification that future multi-spectral crop identification systems will have to be designed to speciate crops. Considering the demand for the technology to speciate crops and predict crop yields, it seems evident that agricultural remote-sensing programs can no longer afford to invest in the fallacy that any multispectral sensor can satisfy objectives of the program.

The literature supports the latter conclusion by the simple fact that only two remote-sensing techniques have demonstrated excellent capability to remotely identify crops: 1) the narrowband multispectral camera system having detection band-passes selected on the basis of spectroradiometry information of the subject material; and 2) the use of the temporal parameter of temporally-registered, multispectral image data. The uniqueness of the two approaches is the fact that they provide specialized tone and texture features that are optimal for speciating a crop using semiautomatic or optical techniques.

379

In conclusion, the objective of future earth resources remote sensing technology should establish the requirement to develop specialized narrowband multispectral sensors that optimally capitalize on the reflectance differences of dynamic and dormant materials. With respect to semiautomatic crop identification, if the proposed sensors can speciate the crop, the image processing software will be able to speciate the crop.

Preceding investigations of this reported effort have demonstrated that crop yield prediction may be feasible using only remotely-sensed data; specifically, temporally-registered, multispectral imagery collected during dynamic periods of the crop's growth cycle. Certainly, as mentioned in preceding sections, the feasibility of crop yield prediction using temporally-registered, multispectral imagery will have to be reverified through future studies. However, since there would be no point in attempting to predict the yield of a crop using multispectral imagery in which the crop could not be identified without ground truth, another conclusion can be extracted from this study. That is, future remote sensing systems, which are dedicated to providing data for crop yield prediction models dependent on multispectral imagery data, will require a

380

narrowband multispectral sensor that is dedicated to speciating the subject crop(s).*

---

*Note: The latter conclusion may seem simplistic and obvious. However, after analyzing current national programs that are attempting to demonstrate the feasibility of yield prediction using broadband multispectral imagery that demonstrates only marginal crop identification accuracies, it becomes apparent that there is a need for such obvious statements so that the remote-sensing scientists can redirect their activities to reach the objectives they may have lost site of.

GENERAL CONCLUSIONS

The primary objective of this entire program was to determine the feasibility of predicting the yield of an agricultural crop using temporally-registered, multispectral imagery. However, during the course of the two year program, several subsidiary investigations were conducted. This section summarizes only the most significant conclusions of the various subsidiary investigations. Elaborate interpretations and conclusions are presented in the body of the report at the conclusion of each subsidiary study's technical discussion.

Conclusions of the Spectroradiometry Survey and the Additive Color Analysis of the Multispectral Photography

1. The set of four narrowband multispectral camera filters, which were selected for discriminating levels of stress within corn, functioned satisfactorily. The selected filters' bandpasses are: 406-480 Nm(blue); 541-581 Nm(green); 620-660 Nm (red); and 810-885 Nm(near-infrared).

2. The multispectral filters were not selected to demonstrate the speciation of corn from other crops. Nevertheless, it was possible to speciate corn by analyzing the multispectral imagery collected during mid-July with additive color analysis techniques.

382

3. The spectral reflectance characteristics of corn was observed to change during the course of the growth cycle. The dynamic spectral reflectance of corn was attributed to changing leaf to soil ratios, changing plant configurations (i.e., leaf shape, tassel, etc), and normal changes in the health of the plants. For this reason, selection of narrowband filters for speciating corn could not be expected to function adequately during the entire growth cycle.

Note: Elaborate interpretations and conclusions of this phase of the program may be referenced on pages 112-113, & 391-408.

Conclusions of the Feasibility Study of Semiautomatic Crop Stress Level Mapping and Agricultural Crop Yield Measurement

1. The feasibility of semiautomatically mapping levels of stress within an agricultural crop, corn, was demonstrated with the digitized, spectrally-registered, multispectral images of three agricultural test sites.

2. When complemented by ground truth information (i.e., yield data/stress level class), the digital stress level maps of the three sites served as crop yield maps. The digital stress level maps generated by image processing the multi-spectral imagery of September or October, were utilized to measure the yield of the three test sites with 96% to 98% accuracies.

3.   The digital stress level maps of each site which were
produced from either September or October's three spectral
images, provided the best data set for measuring the average
yield of each site with consistent excellent accuracies (96% to
98%). Note: Elaborate interpretations and conclusions of this
phase of the program may be referenced on pages 237-242

Conclusions of the Multivariate and Regression Analysis Studies
which were utilized to investigate the Feasibility of Agricul-
tural Crop Yield Prediction using a temporally-registered
multispectral image data set

1.   The feasibility of timely (as early as pollination
phase) and accurate agricultural crop yield prediction (87% of
the actual crop yield data variance was accounted for with a
90% accuracy), of fourteen small corn fields in Central Penn-
sylvania, was demonstrated utilizing a yield prediction equation
that was solely dependent upon remotely-sensed data; specifi-
cally, a digital, temporally-registered, multispectral image
data set.

2.   When compared to the variables defining the time-
variant multispectral image data, the weather variables were
observed to be insignificant predictors of agricultural crop
yield.

384

3. Though the multispectral image data sets, collected on 29 April, 2 July, 16 July, 19 August, 4 September, and 8 October 1974, were analyzed to determine the feasibility of crop yield prediction, the variables defining imagery collected during mid-July and mid-August were prime predictors of agricultural crop yield. Optimal multispectral imaging periods for accurate yield prediction of the same Pennsylvania test sites may actually occupy some subset of the latter significant time interval.

4. Though the two temporally-registered multispectral image data sets of mid-July and mid-August were found to be adequate for timely and accurate yield prediction, the analysis of three or more image data sets, spanning from early July to early September, did increase the amount of yield data variance accounted for to nearly 100%, with a 90% accuracy.

5. The analysis of image data extracted from two or more bands of multispectral imagery registered over time, provided better yield prediction results than when only one band's imagery was utilized. The optimal two-band yield predictor combination was the green/near-infrared image data. The best accuracies in yield prediction occurred when image data from all three spectral images was analyzed.

6. The average density of each of the digital multispectral images of the three large corn agricultural test sites, defined by the variable "GLTX", was highly correlated to weather data. In a regression equation for yield prediction, the GLTX image data could be substituted for weather data with no decrease in the crop yield prediction accuracies.

Note: Elaborate interpretations and conclusions of this phase of the program may be referenced on pages 334 - 337 .

Conclusions of the Feasibility Study of Semiautomatic Crop Identification using Multispectral Imagery dedicated to Crop Stress Detection

1. Semiautomatic crop identification was demonstrated with only an estimated $50 \pm 30\%$ accuracy, using narrowband multispectral image data that was dedicated to detecting levels of stress within corn. Such mediocre crop identification accuracies, achieved with semiautomated image processing techniques, was attributed to the fact that the narrowband multispectral camera system was dedicated to corn stress level detection and not to the speciation of corn.

2. If the feasibility of crop yield prediction using temporally registered, multispectral image data is verified in future studies, futuristic crop yield prediction models will require image data from a narrowband multispectral sensor that is dedicated to speciating the subject crop(s).

386

3. The only known alternative technique for obtaining better semiautomatic crop identification accuracies is the use of temporally-registered multispectral image data. The temporal parameter, in conjunction with a multispectral sensor system dedicated to crop speciation, may be the ideal combination for achieving crop identification accuracies suitable for crop yield prediction models which require temporally-registered, multispectral image data.

Note: Elaborate interpretations and conclusions of this phase of the program may be referenced on pages 378 - 381 .

RECOMMENDATIONS

A program of significant magnitude, spanning an estimated three years, should be developed to verify the demonstrated feasibility of agricultural crop yield prediction using temporally-registered, narrowband multispectral imagery. The proposed study should be directed to demonstrating high accuracies in semiautomatic crop identification and yield prediction using solely, remotely-sensed data.

The first year of the program should be devoted to collecting in situ spectroradiometric data of important agricultural crops and typical background materials during all phases of the respective crop growth cycles. The analysis of the radiometric data would determine the narrowband filters or bandpasses of the detectors for optimally speciating the selected crop(s) during the growth cycle.

During the second year, a calibrated narrowband multispectral sensor, either photographic or electro-optical, operating at the selected crop speciating bandpasses, would be flown regularly at low and high altitudes over geographically and environmentally diverse test sites of the United States. Superlative ground and image truth would supplement the multispectral imagery collection effort.

388

The narrowband multispectral imagery would then be digitally analyzed with an image processing computer system that was dedicated to handle the volume of data collected during the study. Semiautomatic crop identification would be investigated with image processing techniques. The feasibility of crop yield prediction using the temporally-registered, multispectral imagery would be re-investigated using multivariate and regression analysis techniques during the third year of the study.

The proposed study would verify or deny the findings of this reported effort. It would also establish the semiautomatic crop identification and crop yield prediction accuracies that would be potentially obtainable using temporally-registered, narrowband multispectral image data. Additionally, the key phases of the growth cycle for yield prediction and/or crop identification of the economically-important agricultural crops could be established. System effectiveness over geographically-distinct environments could also be measured. Finally, if successful, the investigation would provide valuable information for specifying the design parameters of an operational satellite system that would be dedicated to remote crop identification for crop yield prediction surveys.

389

Appendices A thru G

APPENDIX   A

Spectral Reflectance Characteristics of Corn*

*Appendix A paraphrases a portion of text from an unpublished
manuscript, reference: Grodewald, et al.(20). Permission to
paraphrase the text was granted by the author.

391

Introduction:

A telespectroradiometer was used to obtain in situ reflectance spectra of corn (i.e., Figure A-1), afflicted by varying degrees of stress, and other agricultural crop backgrounds during three phases of the corn growth cycle. This appendix reviews the theory, analysis methodology, interpretations, and conclusions derived from the analysis of the spectral reflectance data.

Seasonal Characteristics of the Spectral Reflectance of Corn

The in situ percent spectral reflectance of corn varies depending on the stage of growth. Though the spectral signature of any type of vegetation is dynamic, it is particularly true of corn because the normal changes transpiring during the corn growth cycle are dramatic compared to other grain crops such as wheat or oats.

During this study, the growth cycle for "normal corn" was conveniently separated into four major divisions:

Spring: The immature plants display a uniform verdant color. A primary factor affecting the spectral signature of an individual plant is the leaf to soil ratio. When considering a particular area of a field, the spectral signature is a function of the plant size and population.

J SCAN # 3125 Grand average - healthy corn - July 1974

PERCENT DIRECTIONAL REFLECTANCE

PDR (%X10)	WVL (NM)	0	10	20	30	40	50	60	70	80	90	100
44	400											
45	412											
50	425											
55	437											
59	450											
63	462											
65	475											
68	487											
72	500											
85	512											
110	525											
137	537											
155	550											
160	562											
154	575											
140	587											
132	600											
128	612											
122	625											
119	637											
116	650											
106	662											
115	675											
148	687											
288	700											
421	712											
587	725											
678	737											
747	750											
780	762											
795	775											
800	787											
808	800											
814	812											
819	825											
823	837											
833	850											
841	862											
837	875											
832	887											
819	900											
818	912											
811	925											
801	937											
773	950											
754	962											
759	975											
772	987											
790	1000											
814	1025											
833	1050											
844	1075											

(%X10) (NM)

Figure A-1. Grand average of all spectral reflectance measurements of healthy, mature corn at the Pennsylvania State University Agronomy Farm in July 1974. Arrows adjacent to the left column indicate the bandpass of the camera filters.

393

Early
Summer: The plants are a verdant color and occupy a large

percentage of the surface area of the field with the

leaf to soil ratio approaching one. The leaves are

long and slender and do not contribute a significant

percent of shadow to the overall reflectance of

the fields.

Mid
Summer: The individual plants are now topped by tassels.

The tassels are initially green but become brown

as the pollen is dispersed. The leaves are becoming

broader and produce a significant amount of shadow

to the overall reflectance of the field.

Autumn: The plants become brown in color. By the time the

corn is harvested, the corn plants are biologically

dead. The spectral signature at this time is affected

to a great extent by the plant's state of dehydration.

The spectral reflectance of stressed corn recorded during
one portion of the normal growth cycle was observed to be spec-
trally similar to normal corn at a different time of the growth
cycle. The most outstanding examples of the latter phenomena
were the extensively blighted corn of mid and late summer.
This corn appeared spectrally similar to normal preharvest corn
in all respects other than potential yield. Corn that displays

394

severe hydrostress later in the growth cycle may have the same spectral characteristics as normal corn at earlier stages of the growth cycle.

Agricultural Crop Stresses, Radiometry, & Filter Selection

One of the primary objectives of this experiment (reference the Preface), was to demonstrate the feasibility of predicting the yield of an agricultural crop, corn, by utilizing the information of a narrowband multispectral data set collected during an entire growth cycle.

The yield of areas within a corn field are a function of the physiological stresses acting within those areas. Since the actual yield figures cannot exist until the crop has been harvested, it is not possible to base the acquisition of reflectance spectra during the growth cycle on any locations of a known yield. Consequently, spectral reflectance measurements were collected in areas where known stresses would positively or negatively affect the localized yield. Finally, the stress-detecting multispectral camera filters were selected from the analysis of the reflectance spectra.

Hydrostress, soil-induced stresses, corn blight, and weed infestations were types of stress that were investigated during this study. The respective spectral signatures of corn affected by these stresses were collected. Certainly, many other stress

factors affecting corn yield must exist but were not observed during the ground monitoring of the various test sites (i.e., insect infestations, root disease, grubs, etc.).

Since the literature of yield prediction studies indicated the significance of an agricultural crop's leaf to soil ratio's relationship to yield measurement, the multispectral filters were required to distinguish vegetation from soil. Fortunately, due to the large difference which exists between the reflectance spectra of soils and vegetation, any set of multispectral filters which occupied portions of the visible and near-infrared spectrum would produce a unique difference between the color signature of bare soil and vegetation in an additive color presentation (refer to text on equipment, page 62).

A Comparison of the Reflectance Spectra of Various Crops

When multispectral photography is utilized, there is a greater concern for relative energy levels rather than absolute reflectance values between imaged materials. Appreciable differences exist between the spectral reflectance of corn compared to other crops at certain wavelengths of the visible spectrum. If a multispectral camera images these spectral regions, the imaged corn and background vegetation will display different relative densities.

In order to differentiate the corn from the other crops
which predominate the Nittany Valley agricultural test sites,
the spectral reflectance of corn was compared to other vege-
tation growing there. This investigation of crop spectral
signatures was conducted using the data collected at the
Pennsylvania State University Agronomy Farm during July and
September of 1974. In addition to healthy mature corn, the
reflectance spectra of soybeans, wheat, pasture grass, oats,
alfalfa, and barley were collected. The latter spectral
reflectance measurements were collected during early September,
with the exception of wheat and oats which were measured in
late July 1974.

The relative percent reflectance of the aforementioned
crops with respect to corn was computed with the following
algorithm:

$$R_r = \frac{(R_c - R_b) \times 100}{R_c}$$

R_r = Relative reflectance ratio (percent)

where: R_c = Percent directional spectral reflectance of corn

R_b = Percent directional spectral reflectance of the
background crop

The relative percent spectral reflectance is tabulated as
a function of wavelength in Table A-1. The spectral bands of
the stress detection filters are annotated on the wavelength
column.

397

BASE SCAN NUMBER IS: 3501 September Corn Grand Average

SCAN # →	2568	2588	2608	3450	3463	3480
WVL	Oats	Wheat –	Barley	Pasture Grass	Soybeans	Alfalfa
400	-23.91	-56.52	-30.43	-30.43	-23.91	-28.26
412	+28.00	+58.00	+32.00	+36.00	-26.00	+34.00
425	+29.62	+55.55	+31.48	+27.77	+29.62	+35.18
437	+31.03	+55.17	+25.86	+27.58	+31.03	+37.93
450	+33.87	+54.83	+24.19	+29.03	+33.87	+43.32
462	+37.87	+54.54	+21.21	+31.81	+36.36	+43.93
475	+40.57	+53.62	+17.39	+33.33	+39.13	+46.37
487	+42.46	+53.42	+16.43	+35.61	+39.72	+49.31
500	+43.42	+52.63	+11.84	+34.21	+40.78	+50.00
512	+44.18	+48.83	+10.46	+32.55	+37.20	+47.67
525	+40.38	+44.23	+12.50	+24.03	+23.07	+38.46
537	-37.70	-40.98	-13.93	+17.21	-9.01	+29.50
550	+36.49	+39.41	+13.86	+15.32	+4.37	+27.00
562	+37.06	+38.46	+11.18	+16.78	+5.59	+28.67
575	+39.28	+38.57	-5.71	+20.00	-11.42	+33.57
587	+42.85	+39.09	+2.25	+25.56	+21.80	+41.35
600	+45.73	+39.53	+8.52	+30.23	+30.23	+46.51
612	+47.65	+41.40	+13.28	+32.61	+35.15	+50.00
625	+49.60	+41.60	+19.20	+35.20	+40.80	+53.60
637	+52.41	+43.54	+22.58	+36.29	+44.35	+57.25
650	+54.40	+46.40	+24.00	+40.00	+48.80	+60.00
662	+57.72	+48.73	+26.01	+45.52	+55.28	+63.41
675	+53.83	+46.21	+22.72	+33.63	+44.69	+54.54
687	+43.04	+39.07	+19.20	+23.17	+22.51	+39.07
700	+25.77	+27.04	+7.40	+2.04	+18.03	+7.40
712	+14.61	+26.02	-18.71	+7.33	+33.94	+19.86
725	-9.26	+28.66	-34.69	+3.87	+36.42	+27.37
737	+9.58	+31.28	+41.41	-0.54	+34.53	+28.39
750	+6.13	+39.51	+43.11	-2.81	+32.66	+28.85
762	+5.04	+30.44	+43.69	-4.57	+30.75	+27.60
775	+5.10	+29.56	+43.34	+4.95	+30.18	+27.55
787	+4.27	+28.09	+42.13	+4.58	+29.31	+27.48
800	+4.54	+26.96	+40.45	+4.54	+29.39	+27.57
812	+4.06	+25.30	+38.55	+4.96	+28.76	+27.49
825	+5.79	+25.40	+38.03	+5.85	+28.08	+26.30
837	+6.31	+24.96	+38.61	+4.69	+27.45	+25.69
850	+7.01	+24.56	+38.30	+4.38	+27.19	+26.16
862	+6.81	+24.20	+37.68	+5.21	+26.66	+25.65
875	+6.64	+23.98	+36.41	+4.76	+26.30	+25.86
887	+4.62	+23.69	+36.70	+4.47	+26.30	+25.14
900	+4.33	+23.84	+35.54	+4.04	+26.73	+27.45

Table A-1. The ratio of percent reflectance of background crops compared to September healthy mature corn. Arrows adjacent to the first column indicate bandpass of the filters utilized.

Weed Infestation

Since the growth of weeds can significantly reduce crop yield, weed control is an important consideration. In order to include the extent of weed infestation in a corn yield prediction model, one of the criteria for the selection of multi-spectral camera filters is that the resultant imagery differentiate corn from resolvable concentrations of weeds.

Common weed types found in the Nittany Valley region included nutsedge, edge grass, and California thistle. The reflectance spectra of these weed types were measured, Figures A-2 to A-5, in early September of 1974 (weed reflectance spectra are included in the appendix of reference (20)).

398

The percent relative reflectance ratios of these three weed species, with respect to the reflectance of healthy, mature corn of September, are tabulated in Table A-2. At least during September, the spectral reflectance of the weeds is less than that of corn in all four spectral bands of the stress detection filters.

Hydrostressed Corn

During the early part of the corn growth cycle, June, the only visible stress is due to excess water accumulating in areas of the corn fields. This hydrostress condition is due to soil permeability and topographic expression. The location of where hydrostress occurs or has been occurring is identifiable by ground conditions and substandard corn plant size. The reflectance spectra of healthy and hydrostressed June corn were recorded (refer to pages 45-49). A comparison of the relative reflectance of hydrostressed June corn to healthy June corn is tabulated in Table A-3.

In comparison to healthy June corn, hydrostressed corn is brighter in the blue, 406-480nm, and red, 620-660nm, bands; nearly equivalent spectral reflectance in the green, 541-581nm band; and a lower reflectance in the near-infrared, 810-885nm band. These spectral reflectance differences are frequently reduced to insignificant levels in the multispectral photography collected during June, due to the amount of soil which is

399

Figure A-2.   Healthy Corn at Field
One, July 1974

Figure A-3.   California
Thistle Weed Infestation
at Field Two, Sept. 1974

Figure A-4.   Severe Corn
Leaf Blight, Field Six,
Sept. 1974

Figure A-5.   Combination Stress
of Blight, Weeds, and Insect
Damage at Field Two, Sept. 1974

400

S3501   RELATIVE REFLECTANCE RATIOS (PERCENT)

BASE SCAN NUMBER IS: 3501 September corn grand average
SCAN #>   3353   3418   3670
  WVL   Edge   California
        Nutsedge   Grass   thistle

WVL	Nutsedge	Edge Grass	California thistle
400	-53.19	-34.04	-27.65
412	-50.98	-31.37	-27.45
425	-49.09	-27.27	-23.63
437	-47.45	-27.11	-23.72
450	-47.61	-26.98	-25.39
462	-46.26	-29.85	-26.86
475	-47.14	-30.00	-27.14
487	-45.94	-31.08	-28.37
500	-44.15	-28.57	-27.27
512	-39.08	-24.13	-25.28
525	-32.38	-16.19	-21.90
537	-26.82	-9.75	-17.88
550	-26.08	-7.97	-16.66
562	-25.69	-8.33	-15.97
575	-25.53	-9.21	-16.31
587	-26.86	-11.94	-17.91
600	-26.92	-14.61	-18.46
612	-27.90	-17.05	-19.37
625	-28.57	-18.25	-20.63
637	-29.60	-20.80	-21.60
650	-33.33	-21.42	-23.80
662	-38.70	-25.00	-25.00
675	-31.57	-21.05	-23.30
687	-23.68	-13.15	-17.10
700	-6.93	+2.04	-4.89
712	-12.53	-3.79	-9.62
725	-22.15	-13.97	-18.49
737	-29.24	-18.41	-23.10
750	-31.62	-20.86	-25.49
762	-32.59	-23.14	-27.40
775	-33.07	-22.41	-27.20
787	-32.92	-21.79	-26.21
800	-31.77	-21.18	-26.17
812	-31.57	-21.35	-26.31
825	-31.45	-22.25	-27.44
837	-31.23	-22.43	-27.71
850	-31.24	-22.18	-27.44
862	-30.96	-21.99	-27.20
875	-30.88	-20.63	-26.40
887	-30.15	-19.19	-25.39
900	-30.44	-18.90	-25.68
912	-31.16	-16.73	-25.39
925	-30.47	-15.67	-26.12
937	-29.69	-12.37	-18.77
950	-28.38	-17.03	-25.32
962	-27.35	-18.52	-27.79
975	-27.90	-18.06	-26.87
987	-28.55	-18.98	-27.10
1000	-29.17	-18.13	-25.92
1025	-28.90	-14.24	-24.20
1050	-31.66	-17.73	-23.12
1075	-25.06	-15.06	-23.33

Note: Arrows adjacent to left column show band pass of the camera filters.

Table A-2. The ratio of percent reflectance of three weed types compared to September healthy, mature corn.

401

BASE SCAN NUMBER IS: 2631 Grand average of June healthy corn.
SCAN #> 2672   June hydrostressed corn.

WVL	
400	+20.45
412	+20.45
425	+19.56
437	+20.83
450	+22.44
462	+22.00
475	+26.00
487	+26.92
500	+25.45
512	+22.22
525	+13.41
537	+4.76
550	+1.63
562	+2.34
575	+7.37
587	+16.51
600	+22.54
612	+26.26
625	+31.91
637	+39.32
650	+44.18
662	+57.50
675	+53.48
687	+34.54
700	+3.88
712	-9.89
725	-18.40
737	-23.05
750	-24.22
762	-24.30
775	-24.78
787	-24.73
800	-24.32
812	-23.85
825	-24.18
837	-24.59
850	-25.00
862	-25.34
875	-25.00
887	-24.55
900	-25.39
912	-25.24
925	-25.04
937	-23.70
950	-22.64
962	-22.33
975	-22.64
987	-22.37
1000	-22.65
1025	-22.98
1050	-21.50
1075	-22.02

Note: Arrows adjacent to left column show band pass of the camera filters.

Table A-3. The ratio of percent reflectance of hydrostressed June corn compared to healthy June corn.

imaged (low leaf to soil ratio). Where the leaf to soil ratio
is 75% or greater, hydrostress can be easily detected due to
the characteristic spectral reflectance differences.

During July, the reflectance spectra of corn, which had
been previously hydrostressed, was measured again. By this
time the original stressed area had all but visually dis-
appeared; that is, the corn appeared more uniform to the eye.
Consequently, identification of hydrostressed areas was rather
difficult. However, the same criteria of topographic expres-
sion, soil permeability, and plant size was utilized. The
July reflectance spectra are included in the appendix of
reference (Grodewald,(20)). Table A-4 illustrates the relative
reflectance of July hydrostressed corn compared to normal July
corn, in which spectral reflectance differences have been
greatly reduced compared to those existing in June (Table A-3).

September Corn Stresses: Weeds, Corn blight, and Hydrostress

By late summer, the stress patterns in corn are virtually
well established. Areas within the corn field which were
hydrostressed in June, now contain plants of inferior size and
vigor compared to normal corn of September. Areas of corn
victimized by Northern Corn Leaf Blight were easily identified.
The severity of this disease varied from minor lesions, accom-
panied by a slight chlorosis, to the condition of almost total

403

S3125   RELATIVE REFLECTANCE RATIOS (PERCENT)

BASE SCAN NUMBER IS: 3125   Grand average of July healthy corn
SCAN #> 3101   July hydrostressed corn.
  WVL

WVL	Ratio
400	+4.54
412	+2.22
425	+.00
437	-1.81
450	-3.38
462	-7.93
475	-9.23
487	-7.35
500	-6.94
512	-3.52
525	+1.81
537	+4.37
550	+6.45
562	+6.25
575	+4.54
587	+2.14
600	+.75
612	-.78
625	-4.09
637	-9.24
650	-12.93
662	-16.98
675	-17.39
687	-12.16
700	-3.47
712	-5.70
725	-11.41
737	-12.97
750	-14.99
762	-16.15
775	-16.60
787	-16.25
800	-16.21
812	-15.47
825	-16.11
837	-16.52
853	-16.92
862	-17.12
875	-16.00
887	-15.38
900	-14.40
912	-15.03
925	-14.18
937	-12.73
950	-13.84
962	-15.11
975	-15.67
987	-15.80
1000	-15.18
1025	-14.61
1050	-15.24
1075	-11.49

Note: Arrows adjacent
to left column show
band pass of the
camera filters.

Table A-4. The ratio of percent reflectance of previously
hydrostressed corn in July compared to healthy July corn.

plant destruction, with the ears of corn being consumed by smut.
The following stress conditions were spectrally measured and
are included in the appendix of reference (Grodewald,(20)).

     --Corn, early stage of Northern Corn Blight

     --Corn, with weedy undergrowth

     --Corn, with advanced Northern Corn Blight

     --Corn, previously hydrostressed

The percent relative reflectance of these stresses compared
to the reflectance of normal, healthy, September corn are listed
in Table A-5. From this data it is evident that all stresses
reduce the reflectance of corn in the 810-885nm band; the most
pronounced reduction being caused by advanced Northern Corn
Blight.

Chlorosis is produced by all stresses (except for early
maturing corn with low degree of infection by Northern Corn
Leaf Blight). This is evidenced by increased reflectance
in the 620-660nm band.

Corn leaf blight reduces the reflectance in the chlorophyll
reflectance band, 541-580nm, while corn with a weed undergrowth
increases in reflectance.

Corn which had been previously hydrostressed in June
exhibits small differences in reflectance compared to normal
September corn. Stressed corn seems to be characterized by only
small spectral reflectance differences in the blue, 406-480nm
band.

BASE SCAN NUMBER IS: 3501 September Corn Grand Average

SCAN #>	3559	3615	3653	3745	3791	3831
					Early Maturing Corn	
WVL	Corn & Weeds	Early Blight	Advanced Blight	Hydro-stress.	Lo.Blight	Adv.Blight
400	-4.34	+6.52	-2.17	+.00	+8.69	+6.52
412	-2.00	+6.00	+.00	-4.00	+4.00	+4.00
425	+.00	+3.70	+1.85	-3.70	+3.70	+1.85
437	-3.44	+5.17	+5.17	-3.44	+1.72	+1.72
450	-4.83	+3.22	+1.61	-3.22	+1.61	+1.61
462	-4.54	+1.51	+3.03	-3.03	+.00	+3.03
475	-5.79	+1.44	+5.79	-2.89	-1.44	+4.34
487	-6.84	+.00	+8.21	-2.73	-1.36	+5.47
500	-3.94	+.00	+9.21	-1.31	-1.31	+6.57
512	+.00	-1.16	+4.65	+.00	-4.65	+.00
525	+6.73	-3.84	-5.76	+.00	-10.57	-11.53
537	+12.29	-4.09	-13.11	+3.27	-13.93	-18.85
550	+14.59	-5.83	-16.78	+4.37	-15.32	-22.62
562	+15.38	-5.59	-14.68	+4.89	-15.38	-21.67
575	+15.71	-3.57	-9.28	+7.14	-12.85	-16.42
587	+14.28	-1.50	+.00	+9.02	-9.77	-9.02
600	+12.40	+1.55	+6.97	+11.62	-6.97	-1.55
612	+10.93	+3.12	+14.06	+12.50	-4.68	+3.90
625	+9.60	+5.60	+21.60	+14.40	-2.40	+10.40
637	+8.06	+7.25	+29.03	+14.51	+.00	+16.93
650	+5.60	+6.40	+32.80	+12.80	-.80	+20.80
662	+.81	+6.50	+40.65	+12.19	+.81	+26.82
675	+2.27	+7.57	+40.15	+9.84	-.75	+25.00
687	+5.96	+5.96	+29.13	+10.59	-3.97	+13.90
700	+16.39	+4.09	-1.22	+4.91	-14.34	-18.44
712	+10.52	-1.75	-19.59	-5.26	-21.05	-35.67
725	+1.29	-8.18	-33.62	-15.08	-26.29	-47.41
737	-4.88	-12.83	-39.96	-20.61	-29.83	-52.80
750	-7.46	-14.75	-41.29	-22.55	-30.67	-53.89
762	-9.14	-15.93	-42.27	-23.81	-31.23	-53.24
775	-9.44	-16.25	-41.79	-23.68	-30.95	-53.40
787	-9.46	-16.18	-40.91	-22.90	-30.99	-52.36
800	-8.93	-15.75	-39.84	-22.12	-30.15	-50.90
812	-9.18	-15.66	-39.90	-22.74	-29.66	-49.84
825	-9.36	-15.60	-39.37	-22.73	-29.27	-49.47
837	-9.54	-15.85	-39.06	-22.61	-29.07	-48.75
850	-9.64	-15.64	-38.45	-22.07	-28.50	-48.09
862	-9.85	-15.65	-38.11	-21.88	-28.26	-47.39
875	-9.53	-15.75	-37.57	-20.95	-28.03	-46.38
887	-9.10	-16.04	-36.84	-20.37	-28.03	-44.94
900	-8.67	-15.75	-36.41	-19.94	-27.89	-43.93
912	-8.95	-15.89	-36.41	-19.79	-27.31	-43.78
925	-8.57	-18.45	-36.91	-20.05	-26.74	-42.15
937	-5.83	-13.70	-31.77	-16.03	-25.21	-38.62
950	-6.70	-14.28	-32.36	-17.20	-24.78	-39.50
962	-7.36	-13.69	-32.25	-17.67	-24.59	-39.17
975	-8.52	-12.50	-31.91	-18.82	-24.11	-37.79
987	-9.57	-13.35	-31.64	-19.01	-23.94	-37.73
1000	-10.35	-14.18	-31.48	-18.43	-24.25	-37.73
1025	-9.69	-12.60	-30.33	-18.00	-23.82	-36.42
1050	-9.73	-16.84	-30.78	-15.52	-25.00	-38.28
1075	-9.07	-14.68	-30.30	-15.62	-23.76	-35.64

Note: Arrows adjacent to left column show band pass of the camera filters.

Table A-5. The ratio of percent reflectance of corn with various categories of stress apparent in September compared to normal healthy September corn.

406

Reflectance of Normal Corn during the Growing Season

The spectral reflectance of normal, healthy corn changes during the growing season. In situ spectral reflectance measurements were collected in early June, the middle of July, and early September of 1974. The relative percent reflectance of June and September corn compared to July corn is tabulated in Table A-6.

Healthy June corn exhibits lower spectral reflectance in all four bands compared to healthy July corn. September's healthy, mature corn reflectance is low in both the green, 541-580nm, and near-infrared, 810-885nm, bands and only slightly greater in reflectance in both the blue, 406-480nm, and red, 620-660nm bands.

407

```
S2631/    NORMAL CORN, JUNE 1974
S3125/    NORMAL CORN, JULY 1974
S3501     NORMAL CORN, SEPT. 1974
ENTER BASE SCAN #

S3125          RELATIVE REFLECTANCE RATIOS (PERCENT)

BASE SCAN NUMBER IS: 3125
SCAN #>  2631     3501
   WVL
```

WVL	2631	3501
400	-2.27	+4.54
412	-4.44	+11.11
425	-9.99	+8.00
437	-14.54	+5.45
450	-18.64	+5.08
462	-22.22	+4.76
475	-24.61	+6.15
487	-25.00	+7.35
500	-25.00	+5.55
512	-27.05	+1.17
525	-26.36	-5.45
537	-24.08	-10.94
550	-21.93	-11.61
562	-20.62	-10.62
575	-21.42	-9.09
587	-22.85	-5.00
600	-23.48	-2.27
612	-23.43	+.00
625	-23.77	+2.45
637	-26.05	+4.20
650	-26.72	+7.75
662	-25.47	+16.03
675	-26.08	+14.78
687	-26.35	+2.02
700	-28.81	-15.27
712	-30.64	-18.76
725	-36.28	-20.95
737	-37.46	-18.43
750	-39.35	-19.27
762	-40.51	-18.71
775	-40.75	-18.74
787	-40.50	-18.12
800	-40.59	-18.31
812	-40.93	-18.42
825	-40.53	-17.82
837	-40.34	-17.25
850	-40.57	-17.88
862	-40.54	-17.95
875	-39.90	-17.32
887	-39.42	-16.82
900	-38.09	-15.50
912	-37.65	-15.40
925	-36.12	-15.16
937	-37.45	-14.35
950	-35.57	-11.25
962	-34.21	-9.94
975	-34.38	-10.40
987	-34.71	-10.75
1000	-35.31	-10.75
1025	-35.99	-11.30
1050	-36.49	-8.76
1075	-34.47	-11.25

Note: Arrows adjacent to left column show band pass of the camera filters.

Table A-6. The ratio of percent reflectance of healthy June corn (first column) and healthy September corn (second column) to healthy July corn.

APPENDIX   B

General Ground Truth**

of

Pennsylvania Agricultural Test Sites 2,4,5 & 6

**For detailed ground truth of sites reference (Grodewald,(20))

Introduction

In support of the agricultural crop yield prediction study, a ground truth program was developed and conducted by personnel of the Pennsylvania State University Agronomy and Remote-Sensing Departments. One division of the ground truth program was the ground monitoring of seven large agricultural test sites dispersed within a one hundred mile radius of the Pennsylvania State University. The test sites were large farms whose chief agricultural crop was feed corn. These seven large farms were monitored from April through October of 1974, during which time detailed ground truth was collected.

From April to October, the seven agricultural test sites were overflown by RADC C-131 aircraft and photographed with a narrowband multispectral camera and two mapping cameras. Different films were used with the mapping cameras to provide supporting image truth for the interpretation of the multispectral photography.

Due to such factors as poor weather, pilot error, cloud cover, etc., complete multispectral photography data sets, spanning from April to October, were only available for Fields 1, 3 & 7. The detailed ground truth for the three sites is presented in a preceding section of this report. However, for the sake of continuity, a review of the general ground truth for Fields 2,4,5 & 6 is presented in this appendix. For a discussion of the detailed ground truth of the latter sites, reference Grodewald(20).

410

Field Two is located seven statute miles southwest of the
Pennsylvania State University, near the town of Seven Stars.
The performance of the corn crop in this field was very poor
and was the first of the seven test sites that was harvested.

Figure B-1: Multispectral Additive Color Rendition of Field Two.
Original black & white multispectral photography
was collected on 19 August 1974 at 1150 EST at an
altitude of 10,000 feet AGL. The 541-581 nm band
is imaged as green; the 620-660 nm band is imaged
as red; and the 810-885 nm band is imaged as blue.
Corn was planted on May 1-5 and harvested on
27 September 1974. Average yield was 59.9 bushels
per acre. In this additive color rendition, corn
appears as a light purple color.

Field Two Yield Data
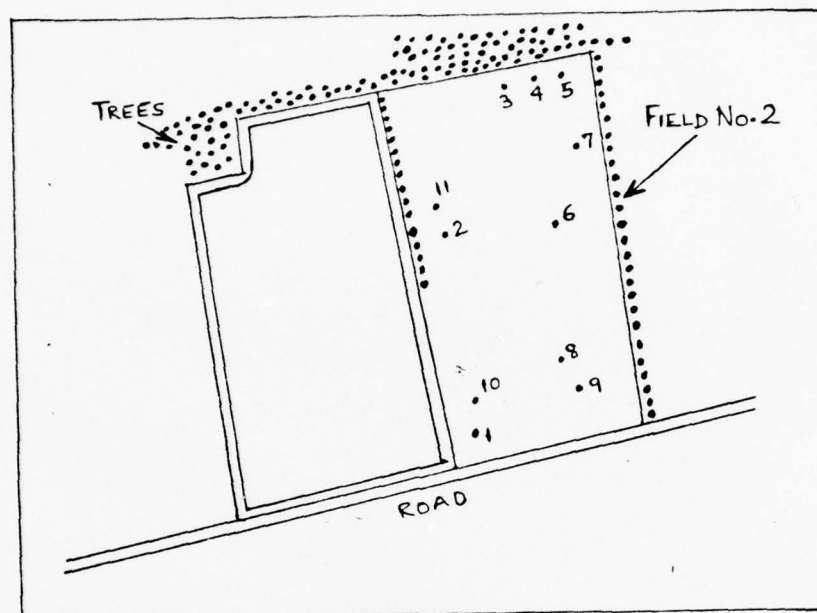
Figure B-2. Locations in Field Two where Yield Measurements
            were collected.

Harvest Date: 27 September 1974

Average Yield: 59.9 bushels/acre

Crops: Corn, Todd M55 and Doeblers 64 varieties

Length of Row Sampled: 27 feet 6 inches

Row Spacing: 38 inches

Sample Area: 1/500 acre

412

Loc.	Gross Wt.	Tare	Net Wt.	# Plants	# Down Plants	# Barren Plants	Wet Wt.	Dry Wt.	Yield Bu/A
1	30 3/4	14	16 3/4	45	0	8	865.6	550.8	90.1
2	29 1/2	14	15 1/2	50	1	15	987.7	603.6	80.1
3	21 3/4	14	7 3/4	36	1	20	999.5	624.6	40.9
4	18 1/2	14	4 1/2	51	3	33	605.0	391.4	24.6
5	15 1/2	14	1 1/2	46	2	39	444.5	274.3	7.8
6	29 1/2	14	15 1/2	41	3	10	1135.9	688.1	79.4
7	29 1/4	14	15 1/4	52	6	14	812.4	506.8	80.4
8	18	14	4	31	1	19	776.6	473.3	20.6
9	32	14	18	46	0	4	1046.8	625.3	90.9
10	27	14	13	48	1	16	949.9	587.3	67.9
11	27 1/2	14	13 1/2	46	2	12	756.5	502.3	75.8

Table B-1. Yield measurements at eleven locations in Field Two.

413

Field Four is located twenty-four statute miles northeast
of the Pennsylvania State University, near the town of Lamar,
Pennsylvania. Field Four was notable for the uniformity of
corn growth throughout the corn growing season.

Figure B-3. Multispectral Additive Color Rendition of Field Four.
Original black & white multispectral photography was
collected on 16 July 1974 at 1138 EST at an altitude
of 10,000 feet AGL. The 541-581 nm band is imaged
as green; the 620-660 nm band is imaged as red; and
the 810-885 nm band is imaged as blue. Exact
planting date is unknown, approximately mid-May.
Field Four was harvested on 1 November 1974. Average
yield was 137.8 bushels per acre.

Field Four Yield Data
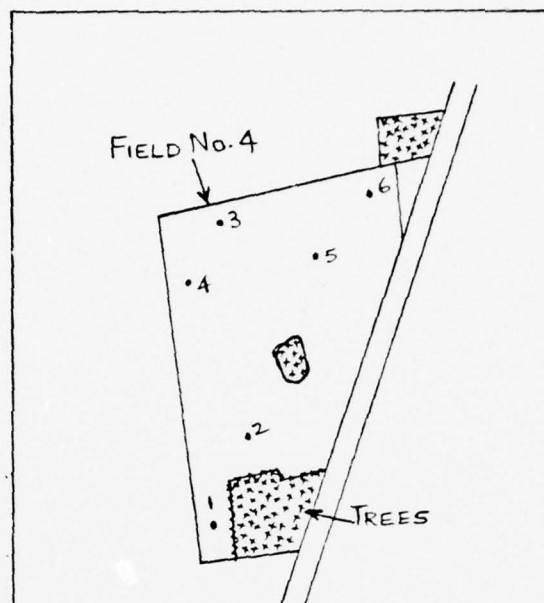
Figure B-4. Locations in Field Four where Yield Measurements were collected.

Harvest Date: 1 November 1974

Average Yield: 137.8 bushels per acre

Crops: Corn (unknown variety)

Length of Row Sampled: 34 feet 10 inches

Row Spacing: 30 inches

Sample Area: 1/500 acre

Loc.	Gross Wt.	Tare	Net Wt.	# Plants	Down Plants	Barren Plants	Wet Wt.	Dry Wt.	Yield Bu/A
1	44 3/4	14	30 3/4	52	1	2	1051	588	143.1
2	42 1/4	14	28 1/4	50	0	5	1106	631	136.3
3	42 1/2	14	28 1/2	54	1	5	1091	656	145.0
4	42	14	28	47	1	1	978	553	133.8
5	44 1/4	14	30 1/4	48	1	1	1100	618	143.8
6	40 1/4	14	26 1/4	47	0	3	988	552	124.5

Table B-2. Yield measurements at six locations in Field Four.

415

Field Five is located one statute mile northeast of
Lockhaven, Pennsylvania, on the east bank of the West Branch
of the Susquehana River.
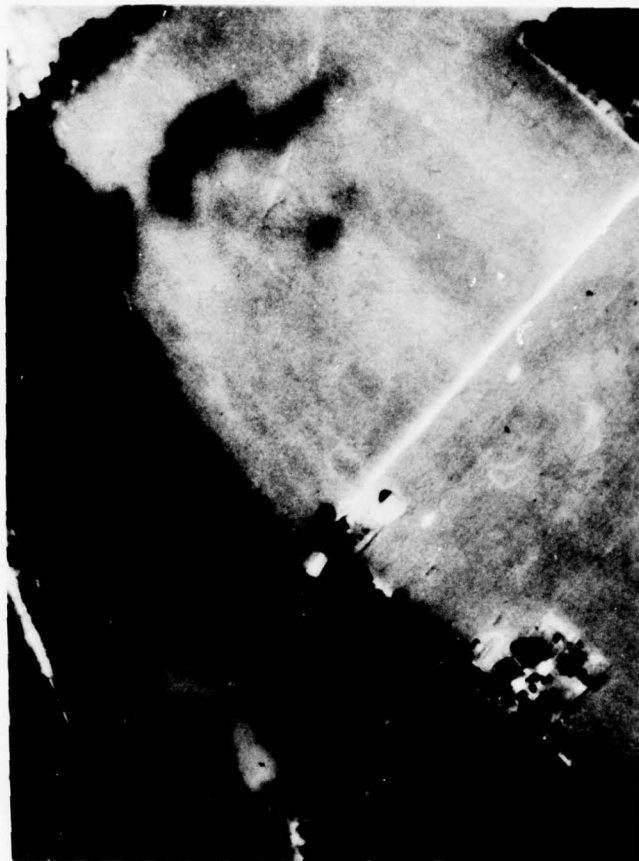
Figure B-5: Multispectral Additive Color Rendition of Field Five.
Original black & white multispectral photography
was collected on 16 September 1974 at 1319 EST at
an altitude of 6,000 feet AGL. The 541-581 nm band
is imaged as green; the 620-660 nm band is imaged as
red; and the 810-885 nm band is imaged as blue.
Corn was planted on May 12th and harvested on
3 November 1974. Average yield was 144.8 bushels
per acre. Light blue spot in field is a weed
infestation. When viewed on the additive color
viewer screen the image of the corn field is a
unique purple color in this rendition.

Field Five Yield Data
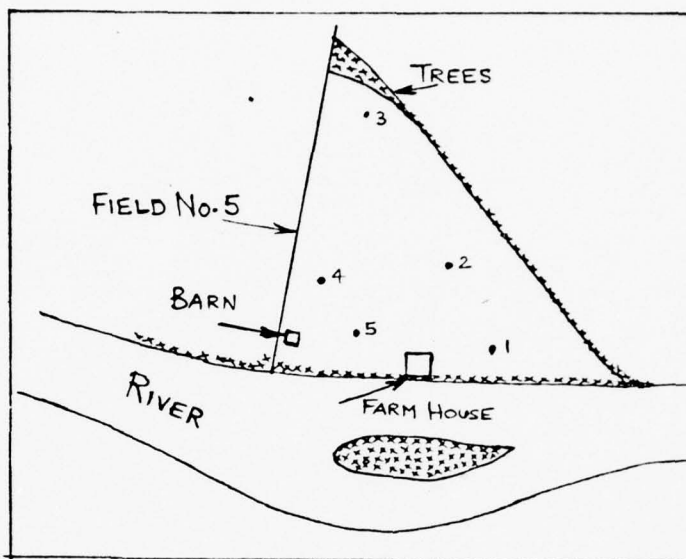
Figure B-6. Locations in Field Five where Yield Measurements were collected.

Harvest Date: 3 November 1974

Average Yield: 144.8 bushels per acre

Crops: Corn, Doeblers 64 varieties

Length of Row Sampled: 34 feet 10 inches

Row Spacing: 30 inches

Sample Area: 1/500 acre

Loc.	Gross Wt.	Tare	Net Wt.	# Plants	# Down Plants	# Barren Plants	Wet Wt.	Dry Wt.	Yield Bu/A
1	50 1/4	14	36 1/4	58	0	5	1132.9	693.3	187.5
2	36 1/4	14	22 1/4	42	0	3	945.2	535.4	106.5
3	38 3/4	14	24 3/4	40	0	5	1168.6	706.6	126.5
4	41	14	27	38	0	2	1229.1	739.0	137.2
5	45 1/4	14	31 1/4	46	0	2	1181.2	727.6	162.7

Table B-3. Yield Measurements at five locations in Field Five.

417

Field Six is located fifteen statute miles directly east of Williamsport, Pennsylvania near the town of Muncy. It is approximately two miles east of the West Branch of the Susquehana River. Three separate fields have been grouped together into Field Six. These fields were managed very well and reported the greatest yield figures at harvest time.
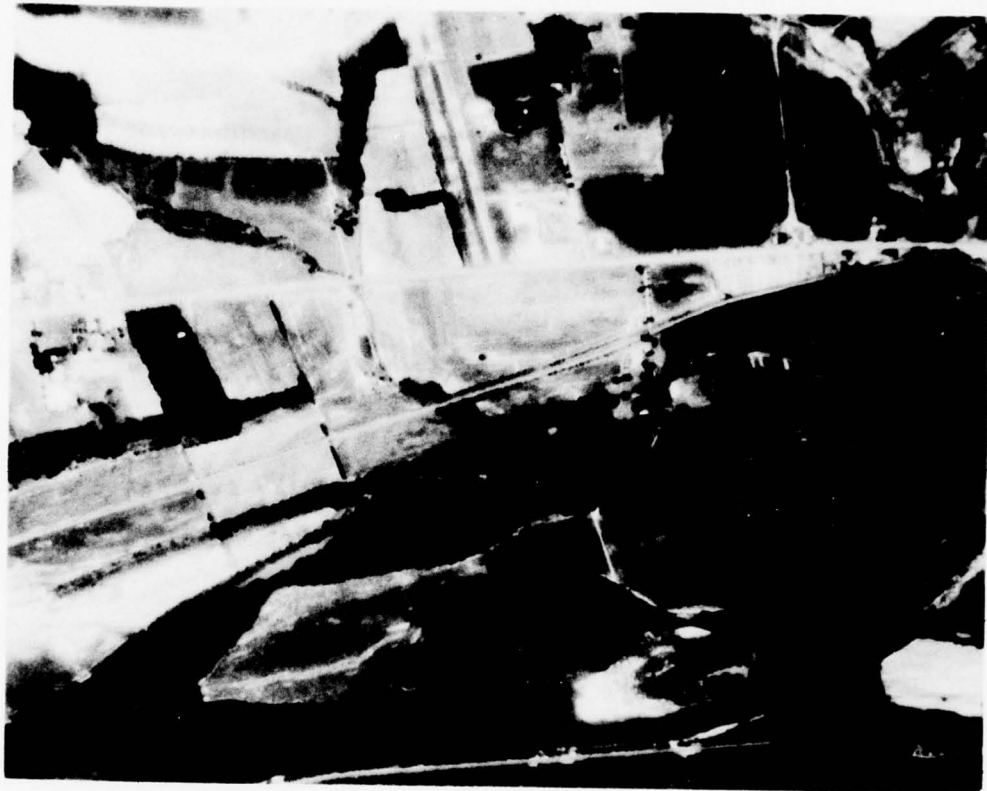
Figure B-7: Multispectral Additive-Color Rendition of Field Six. Original black & white multispectral photography was collected on 12 July 1974 at 1102 EST at an altitude of 25,000 feet AGL. The 541-581 nm band is imaged as green; the 620-660 nm band is imaged as red; and the 810-885 nm band is imaged as blue. On the additive color viewer the color signatures of agricultural crops remained constant for imagery collected from 6,000 to 25,000 feet AGL.

Field Six Yield Data
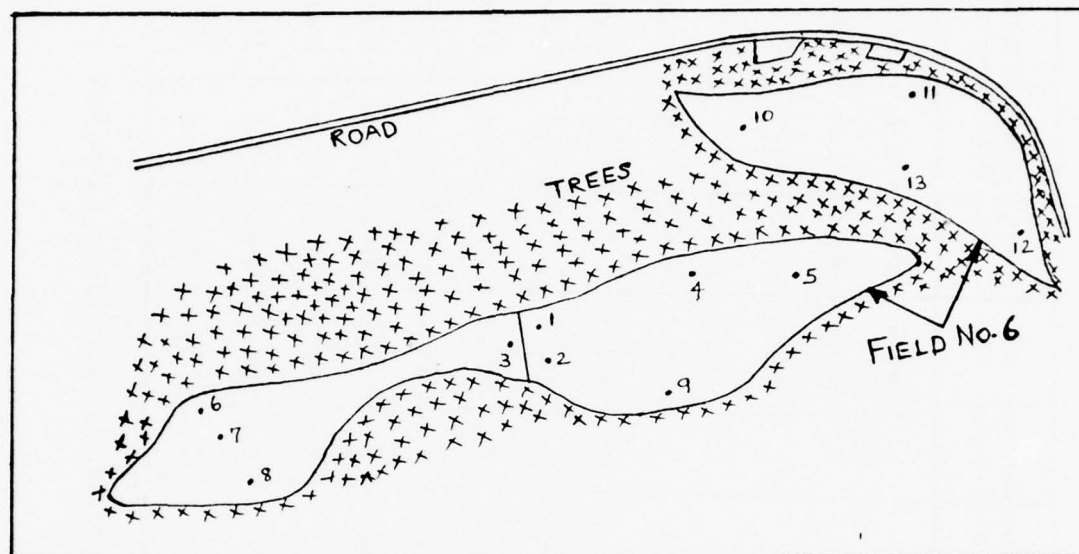
Figure B-8.   Locations in Field Six where Yield Measurements
                were collected.

Harvest Date: Late October 1974 (exact date not

                            recorded)

Average Yield: 158.3 bushels per acre

Crops: Corn, Pioneer 3518 variety

Length of Row Sampled: 26 feet 2 inches

Row Spacing: 40 inches

Sample Area: 1/500 acre

Loc.	Gross Wt.	Tare	Net Wt.	# Plants	# Down Plants	# Barren Plants	Wet Wt.	Dry Wt.	Yield Bu/A	Pioneer Variety
1	48 1/2	14	34 1/2	49	0	1	1088.6	637.8	170.0	3518
2	47	14	33	43	0	1	960.8	567.6	164.8	3784
3	41	14	27	43	1	1	947.2	620.3	149.5	3518
4	45 1/2	14	31 1/2	43	0	0	1083.6	665.8	163.6	3518
5	48 3/4	14	34 3/4	48	0	0	1085.3	640.2	173.3	3518
6	39 1/4	14	25 1/4	37	0	1	947.5	612.5	138.0	3784
7	49 3/4	14	26 3/4	44	0	4	874.0	583.6	151.0	3784
8	40 1/4	14	26 1/4	47	—	1	807.8	538.5	147.9	3784
9	46 1/4	14	32 1/4	48	—	1	1076.0	655.6	166.1	3518
10	37 1/2	14	23 1/2	41	0	4	1070.2	658.9	122.3	3518
11	45 1/2	14	32 1/2	39	0	0	1118.9	678.3	161.4	3518
12	41 1/2	14	27 1/2	42	0	0	1086.0	668.4	143.1	3518
13	39 1/2	14	25 1/2	42	0	3	904.8	558.7	133.2	3518

Table B-4. Yield Measurements at thirteen locations in Site Six

APPENDIX C

Background Mathematics of the Registration Algorithm*

*Note that this Appendix describing the mathematics of the image
registration algorithm was virtually copied from the RADC
Technical Report, entitled Spectral Analysis (Zanon,57) with
the verbal permission of the RADC Project Engineer.

421

General Description:

The purpose of this routine is to register one or more
images with respect to a reference image. This is accomplished
by a combination of rotations, translations and/or scale changes
performed on the image to bring it into register with the refer-
ence image. The method is based on aligning user-specified
control points in both images or by means of user-computed
transformation parameters.

The amount of rotation, translation and scale change is
determined by the formula $y = k^{-1} (Ax+b)$, where x is a vector
of the column and row position of a picture element in the new
image, y is a vector of the column and row position of a pic-
ture element in the original image, A is a two-by-two matrix;
b is a constant vector, and k is a computational scale value.
The following discussion, which consists of a descrip-
tion of the mathematics required for the registration algorithm,
may be ignored by the user without hindering his ability to use
this image registration option. The last paragraph in the
general description may be of additional interest to the user
if he only intends to rotate an image a specific amount.

A new image X can be created by transforming image Y,
making use of the following relationships:

$$y = A\underline{x} + \underline{b}$$

$$\underline{x} = B\underline{y} + \underline{c}$$

where A and B are 2 x 2 matrices $(B = A^{-1})$, and x and y are 2-
dimensional column vectors of Cartesian coordinate values (image
picture element position:

$\binom{column}{row}$, and $\underline{b}$ and $\underline{c}$ are constant vectors $(\underline{c} = -A^{-1} \underline{b})$.

Although the latter equation describes the transformation of
image Y to yield image X, it is the former equation that is
used to create the image X. The grey value to be assigned to
the picture element at x in X is the same as the grey value
already assigned to the picture element y in Y. In the event
that the y vector does not have integer components, a weighted
average of the grey values near y is the assignment for the
grey value of $\underline{x}$.

422

The matrix A and constant $\underline{b}$ may be determined by control points selected on the image Y and a reference image Z. Because the transformation will yield image X, which is registered with respect to Y, a point on image Z will be identically positioned on image X. Thus, the control points selected on image Z are treated as known points on the image X which is to be created.

The matrix A and constant $\underline{b}$ for image translation are uniquely determined by one control point on each of Z and Y; A and $\underline{b}$ for image rotation with translation and uniform scale change (i.e., same, if any, scale adjustment in two independent directions) are uniquely determined by two control points on each of Z and Y; A and $\underline{b}$ for image rotation with translation and non-uniform scale changes (i.e., possibly different scale adjustments in two independent directions) are uniquely determined by three control points on each of Z and Y. The three possible transformations will be referred to as T, R_1, and R_2 respectively, which stand for "TRANSLATION," "ROTATION (SIMPLE SCALE)," and "ROTATION (INDEP. SCALES)."

Notation:

Let

(1) $\underline{x}^i = \begin{matrix} x_1^i \\ x_2^i \end{matrix} = \binom{\text{column value}}{\text{row value}}$ (i-th control point on the reference image Z and hence on the output image X.)

and

$\underline{y}^i = \begin{matrix} y_1^i \\ y_2^i \end{matrix}$ (corresponding i-th control point on the input image Y)

for $0 \leq i \leq N$, where $N + 1$ is the number of control points. Note that the superscript zero simply designates the initially-selected control point on the respective image.

(2) $\Delta x_j^i = x_j^i - x_j^0$

$$j = 1, 2 \text{ and } 0 \le i < N$$

$\Delta y_j^i = y_j^i - y_j^0$

The matrix equation $\underline{y} = A\underline{x} + \underline{b}$ is modified to $\underline{y} = k^{-1}$ $(A\underline{x} + \underline{b})$ where k is a constant used for computational purposes (the latter A and $\underline{b}$ are actually the former A and $\underline{b}$ multiplied by the factor k).

The methods to calculate the matrix A, the vector $\underline{b}$ and the constant k follow:

Transformation T (N = 0):

$A = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$k = 1$

$\underline{b} = \underline{y}^0 - \underline{x}^0$

Transformation R_1 (N = 1):

$A = \begin{matrix} a_{11} & a_{12} \\ -a_{12} & a_{11} \end{matrix}$

where

$a_{11} = \Delta x_1^1 \quad \Delta y_1^1 + \Delta x_2^1 \quad \Delta y_2^1$

$a_{12} = -\Delta x_1^1 \quad \Delta y_2^1 \quad \Delta x_2^1 \quad \Delta y_1^1$

$k = (\Delta x_1^1)^2 + (\Delta x_2^1)^2$

$\underline{b} = k\underline{y}^0 - A\underline{x}^0$

424

Transformation R_2 ($N = 2$):

$$A = \begin{matrix} \Delta y_1^1 & \Delta y_1^2 & \Delta x_2^2 & -\Delta x_1^2 \\ \\ \Delta y_2^1 & \Delta y_2^2 & -\Delta x_2^1 & \Delta x_1^1 \end{matrix}$$

$$k = x_1^1 \, \Delta x_2^2 \, - \, \Delta x_1^2 \, \Delta x_2^1$$

$$\underline{b} = k\underline{y}^0 \, - \, A\underline{x}^0$$

Overdetermined Set of Control Points

Since the matrix A and vector $\underline{b}$ are uniquely determined (modulo the constant k) in the description above, any error in selecting the control points will be reflected throughout the image X. To correct for random observational error, additional control points need to be selected. This yields an overdetermined system of equations, i.e., there may be no exact solution since there are more equations than unknowns. Thus, for any selection for the matrix A and vector $\underline{b}$, if the control points are tested in the equation, an error vector $\underline{\varepsilon}^i$ may be present, i.e.,

$$\underline{\varepsilon}^i = \underline{y}^i - k^{-1} (A\underline{x}^i + b), \; 0 \le i \le N$$

What is desired is, of course, that $\underline{\varepsilon}^i = \binom{0}{0}$, $0 \le i \le N$, but in general, this cannot be expected. A cumulative error is defined to be

$$\rho = \sum_{i=0}^{N} \underline{\varepsilon}^i \, . \, \underline{\varepsilon}^i$$

where . denotes dot product. Values for matrix A and vector $\underline{b}$ are desired such that the error, ρ, is minimized. The following describes the appropriate A, $\underline{b}$, and k which minimize ρ. First, some additional notation is required:

(1) $\quad \bar{\underline{x}} = \dfrac{1}{N+1} \displaystyle\sum_{i=0}^{N} \underline{x}^i , \ \bar{\underline{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}$

$\quad\quad \bar{\underline{y}} = \dfrac{1}{N+1} \displaystyle\sum_{i=0}^{N} \underline{y}^i , \ \bar{\underline{y}} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}$

(2) $\quad \bar{\triangle} x_j^i = x_j^i - \bar{x}_j$

$\quad\quad \bar{\triangle} y_j^i = y_j^i - \bar{y}_j$ $\qquad\quad j = 1, 2 \, ; \ 0 \leq i \leq N$

Transformation T $(N > 0)$:

$\quad A = \begin{pmatrix} N+1 & 0 \\ 0 & N+1 \end{pmatrix}$

$\quad k = N+1$

$\quad b = \displaystyle\sum_{i=0}^{N} (\underline{y}^i - \underline{x}^i)$

Transformation R_1 $(N > 1)$:

$\quad A = \begin{pmatrix} a_{11} & a_{12} \\ -a_{12} & a_{11} \end{pmatrix}$

where

$\quad a_{11} = \displaystyle\sum_{i=0}^{N} (\bar{\triangle} y_1^i \ \bar{\triangle} x_1^i + \bar{\triangle} y_2^i \ \bar{\triangle} x_2^i)$

$\quad a_{12} = \displaystyle\sum_{i=0}^{N} (\bar{\triangle} y_1^i \ \bar{\triangle} x_2^i - \bar{\triangle} y_2^i \ \bar{\triangle} x_1^i$

426

$$k = \sum_{i=0}^{N} \left[(\overline{\triangle} x_1^i)^2 + (\overline{\triangle} x_2^i)^2 \right]$$

$$b = k\overline{y} - A\underline{\overline{x}}$$

Transformation R_2 ($N > 2$):

$$A = A_1 \ A_2$$

where

$$A_1 = \begin{pmatrix} \sum \overline{\triangle} y_1^i \ \overline{\triangle} x_1^i & \sum \overline{\triangle} y_1^i \ \overline{\triangle} x_2^i \\ \sum \overline{\triangle} y_2^i \ \overline{\triangle} x_1^i & \sum \overline{\triangle} y_2^i \ \overline{\triangle} x_2^i \end{pmatrix}$$

$$A_2 = \begin{pmatrix} \sum (\overline{\triangle} x_2^i)^2 & - \sum \overline{\triangle} x_1^i \ \overline{\triangle} x_2^i \\ -\sum \overline{\triangle} x_1^i \ \overline{\triangle} x_2^i & \sum (\overline{\triangle} x_1^i)^2 \end{pmatrix}$$

$$k = \left[\sum (\overline{\triangle} x_1^i)^2 \right] \left[\sum (\overline{\triangle} x_2^i)^2 \right] - \left[\sum \overline{\triangle} x_1^i \ \overline{\triangle} x_2^i \right]^2$$

$$b = k\overline{y} - A\underline{\overline{x}}$$

Note: All the summation symbols are to have the limits $\sum_{i=0}^{N}$.

427

If the user knows that the image is to be rotated about a point $\begin{pmatrix} c \\ r \end{pmatrix} = \begin{pmatrix} \text{column} \\ \text{row} \end{pmatrix}$ in a counter-clockwise direction through an angle of ψ, then the appropriate formula is:

$$\underline{y} = k^{-1}\left[\begin{pmatrix} k\cos\psi & -k\sin\psi \\ k\sin\psi & k\cos\psi \end{pmatrix} \underline{x} + \begin{pmatrix} k(c - c \cdot \cos\psi + r \cdot \sin\psi) \\ k(r - c \cdot \sin\psi - r \cdot \cos\psi) \end{pmatrix} \right]$$

For the specific case when $\begin{pmatrix} c \\ r \end{pmatrix} = \begin{pmatrix} 100. \\ 50. \end{pmatrix}$, $\psi = 2^{\circ}$,

$$\underline{y} = 10^{-5}\left[\begin{pmatrix} 99939 & -3490 \\ 3490 & 99939 \end{pmatrix} \underline{x} + \begin{pmatrix} 180589 \\ -345949 \end{pmatrix} \right]$$

would accomplish the desired rotation.

428

APPENDIX D

Dialogue of FACTAN Computer Library Program

Utilized For

Principal Component Analysis**

**Reference Honeywell User Applications Manual (25)

Introduction

This appendix presents the computer dialogue and computations from one of twenty principal component analyses performed using the FACTAN library routine. The purpose of this particular principal component analysis was to investigate the relative significance of the three imagery-extracted variables, YSAX, DIFX, AND GLTX. The variables' data were extracted from only the green band multispectral images collected during the months of April through October 1974. The following dialogue has been extracted from a computer printout.

0112001

RADC R&D TSS GCOS-GU3 03/03/76 AT 15.766 CHANNEL 3200

LOGON ID-BRAZIL: 62441038RADC
PASSWORD--
XXXXXXXXXXXXX
0 BLOCKS FILE SPACE AVAILABLE

SYSTEM? FORTRAN LIB FACTAN
READY
*RUN
INPUT NUMBER OF TEST AND NUMBER OF SUBJECTS*
=11,18

                    *Program is asking for the number of
                     data observations and the number of
                     variables to be analyzed.
INPUT TEST SCORES FOR SUBJECT
=240 143 135 105 115 176 31 6 13 4 -4 18 209 137 122
101 119 158

(For better understanding, the numbers are replaced by
variable identities.).

INPUT TEST SCORES (IMAGE VARIABLE DATA) FOR CORN MINIFIELD 1
= YSAG YSAG YSAG DIFG DIFG...DIFG GLTG GLTG...GLTG
  MAY   JY2   OCT   MAY JY2   OCT   MAY   JY2   OCT

    ...(TOTAL OF EIGHTEEN VARIABLES'S DATA)

430

((Repeat this input of data until the eleven sets of variable
  data observations is entered))
                      - - - - - - - - - - -
                FACTAN Prints out Correlation Matrix
                Dimensions of Correlation Matrix (18 x 18)

                      - - - - - - - - - - -

            (FACTAN prints out the variance accounted for by
             each loading or principal component followed by
             a listing of that component's factor loadings)

∅.38.......E∅∅= VARIANCE ACCOUNTED FOR BY LOADING 1
FACTOR LOADING   1   IS   (loadings shortened to
                                    two place significance)
∅.55 E∅∅ ∅.76 E∅∅ ∅.87 E∅∅ ......lists eighteen factor
                                    loadings for each of five
                                    determined components
                    _____

                FACTAN performs orthogonal rotation of
                principal components using Varimax Criteria

                    _____

INPUT ERROR CRITERION
=.∅∅1   * (user inputs level at which iterative Varimax
                    Criterion algorithm should terminate)

THE VARIMAX CRITERION =
∅.156∅4695E ∅3

THE VARIMAX CRITERION =
∅.16635∅52E ∅3

            FACTAN CONTINUES VARIMAX CRITERION CALCULATION
            UNTIL INPUT LEVEL OF ERROR CRITERION CANNOT BE
            EXCEEDED, THEN FACTAN PRINTS OUT THE FINAL
            ROTATED LOADINGS FOR EACH OF FIVE PRINCIPAL
            COMPONENTS.

                      - - - - - - - - - - -

THE ROTATED LOADINGS ARE

-.192∅9572E-∅1 ∅.69........E∅∅ ...FACTAN lists eighteen
                                    factor loadings for each of
                                    the five rotated principal
                                    components
*DONE
SYSTEM? BYE

                    _____

                    FACTAN TERMINATES

                              431

APPENDIX E

Syntheses of Important Multivariate Analyses

Utilizing the

FACTAN Principal Component Analysis Program

432

Variables' Definitions

Image-Extracted Variables:

YSAX - Mean digital grey level of the 72 pixel area defining the small minifields on a specific multispectral digital image (Ref: Page 187).

    YSAG - YSAX variable, green image (Month to be specified)
    YSAR - YSAX variable, red image (Month to be specified)
    YSIR - YSAX variable, infrared image (Month to be specified)

GLTX - Mean digital grey level of each entire imaged corn field in which the corn minifields are contained (Ref: Page 187).

    GLTG - GLTX variable, green image (Month to be specified)
    GLTR - GLTX variable, red image (Month to be specified)
    GLTIR - GLTX variable, infrared image (Month to be specified)

DIFX - Actual numeric difference in grey level (nonabsolute) between YSAX and GLTX on a stipulated multispectral digital image (Ref: Page 187).

    DIFG - DIFX variable, green image (Month to be specified)
    DIFR - DIFX variable, red image (Month to be specified)
    DIFIR - DIFX variable, infrared image (Month to be specified)

Weather Variables:

AvT - Average Temperature for a given month at the given test site.
AvHT - Average High Temperature for given month at the given test site.
AvLT - Average Low Temperature for a given month at the given test site.
AvH - Average Humidity for a given month at the given test site.
AvRn - Average Rainfall for a given month at the given test site.

Ground Truth Variables:

YSA - Accurate yield of a corn minifield.
YLT - Total field yield average (bushels/acre) of the large test sites.

Abbreviations for the Months Imagery Data

M or MAY - May   A or AUG - August 19
J2 or JY2 - July 2   S or SEPT - September 4
J16 or JY16 - July 16   O or OCT - October 17

FACTAN PRINCIPAL COMPONENT ANALYSIS

TITLE: Red Band Analysis

Variables Analyzed: YSAR May thru October
                           DIFR May thru October
                           GLTR May thru October

Notes: In this analysis, each month's three image-extracted
variables were considered as independent variables such that a
total of eighteen variables were principal components analyzed.
Data for each of the 18 variables was acquired at a total of
11 small corn fields in Central Pennsylvania.

OBJECTIVE: Compared to the green and infrared band analyses,
this analysis determined which months' image variables were pre-
dominant in the accounting for the total data variance. The
rotated components would be interpreted to determine the sig-
nificant image data and months as a function of spectral band.

Number of Principal Components Calculated: 5

Variance (V) accounted for by each Component:

1	2	3	4	5
V = 46%	V = 19%	V = 13%	V = 9.8%	V = 5.9%

Prime Variables describing Principal Component (Factor Loading $\geq$
   0.85).

1	Loading	2	Loading	3	Loading
GLTR JY16	0.99	DIFR SEPT	0.97	GLTR OCT	0.99
GLTR JY2	0.98	YSAR SEPT	0.96	YSAR OCT	0.90
GLTR AUG	0.98	DIFR AUG	0.86		
GLTR SEPT	0.97				
YSAR AUG	0.93	4	Loading	5	Loading
YSAR JY16	0.88				
DIFIR OCT	0.86	DIFR JY2	0.88	DIFR OCT	0.87
YSAR JY2	0.85				

434

FACTAN PRINCIPAL COMPONENT ANALYSIS

TITLE: GREEN BAND ANALYSIS

VARIABLES ANALYZED: YSAG   May thru October
                      DIFG   May thru October
                      GLTG   May thru October

Notes: In this analysis, each month's three image-extracted
variables were considered as independent variables such that a
total of eighteen variables were principal component analyzed.
Data for each of the 18 variables was acquired at a total of 11
small corn fields in Central Pennsylvania.

OBJECTIVE: Compared to the red and infrared band analyses, this
analysis determined which months' image variables were pre-
dominant in the accounting for the total data variance. The
computed rotated components would be interpreted to determine
the significant image data and months as a function of spectral
band.

Number of Principal Components Calculated: 5

Variance (V) accounted for by each Principal Component:

1	2	3	4	5
V = 38%	V = 24%	V = 13%	V = 10%	V = 7%

Prime Variables describing Principal Component (Factor Loading $\geq$
0.85).

1	Loading		2	Loading		3	Loading
GLTG AUG	0.98		GLTG MAY	0.99		DIFG JY2	0.90
YSAG AUG	0.97		GLTG OCT	0.98		DIFG JY16	0.88
GLTG JY2	0.96		YSAG OCT	0.91			
GLTG JY16	0.90		GLTG SEPT	0.87			

4	Loading		4	Loading
DIFG SEPT	0.98		DIFG MAY	0.89
YSAG SEPT	0.87			

435

FACTAN PRINCIPAL COMPONENT ANALYSIS

TITLE: Near-Infrared Band Analysis

Variables Analyzed: YSIR   May thru October
                      DIFIR   May thru October
                      GLTIR   May thru October

Notes: In this analysis, each month's three image-extracted variables were considered as independent variables such that a total of eighteen variables were principal component analyzed. Data for each of the 18 variables was acquired at 11 small corn fields in Central Pennsylvania.

OBJECTIVE: Compared to the green and red band analyses, this analysis determined which months' image variables were predominant in the accounting for the total data variance. The computed rotated components would be interpreted for determining key specific image variables and months as a function of spectral band.

Number of Principal Components Calculated: 5

Variance (V) accounted for by each Principal Component:

1	2	3	4	5
V = 41%	V = 23%	V = 16%	V = 8%	V = 6%

Prime Variables describing the Principal Components (Factor Loading $\geq$ 0.85).

1	Loading	2	Loading	3	Loading
GLTIR SEPT	0.99	DIFIR JY2	0.90	GLTIR MAY	0.94
GLTIR OCT	0.97			GLTIR AUG	0.90
GLTIR JY16	0.90				
YSIR OCT	0.87	4	Loading	5	Loading
		DIFIR MAY	0.92	DIFIR AUG	0.95

436

FACTAN PRINCIPAL COMPONENT ANALYSIS

TITLE: May Multivariate Analysis

Variables Analyzed:

AvT	YSAG	GLTG
AvHT	YSAR	GLTR
AvLT	YSIR	GLTIR
AvH	DIFG	
AvRn	DIFR	
	DIFIR	

Notes: In this analysis, May's image-extracted variables and weather variables were principal component analyzed. A total of 14 variables' data were acquired at a total of 15 small corn fields in Central Pennsylvania.

OBJECTIVE: The results of this analysis were considered with the analyses of the other month's data to determine if the principal components were defined by predominant image variables or weather variables.

Number of Principal Components Calculated: 2

Variance (V) accounted for by each component:

1	2
V = 52%	V = 32%

Prime variables describing the Principal Component (Factor Loading $\geq$ 0.85).

1	Loadings	2	Loadings
AvT	0.99	DIFG	0.93
AvHT	0.99	DIFR	0.86
AvLT	0.99	DIFIR	0.85
AvH	0.99		
AvRn	0.99		
GLTR	0.97		

Comments:

Correlation Matrix:
- Weather data variables were highly intercorrelated, expected
- Weather data variables were highly correlated to GLTR, unexpected

437

FACTAN PRINCIPAL COMPONENT ANALYSIS

TITLE: July 2 Multivariate Analysis

Variables Analyzed: (Reference Variable Definitions, Appendix E, Pg. 433).

AvT	YSAG	GLTG
AvHT	YSAR	GLTR
AvLT	YSIR	GLTIR
AvH	DIFG	
AvRn	DIFR	
	DIFIR	

Notes: In this analysis, July 2's image-extracted variables and weather variables were principal component analyzed. A total of 14 variables' data were acquired at a total of 15 corn minifields in Central Pennsylvania.

OBJECTIVE: The results of this analysis were considered with the analyses of the other month's data to determine if the principal components were defined by predominant image variables or weather variables.

Number of Principal Components calculated: 3

Variance (V) accounted for by each component:

1	2	3
V = 60%	V = 20%	V = 13%

Prime variables describing the Principal Component (Factor Loading $\geq$ 0.85).

1	Loadings	2	Loadings	3	Loadings
AvT	0.99	DIFG	0.83	DIFIR	0.97
AvHT	0.99	DIFR	0.90	YSIR	0.88
AvLT	0.99				
AvH	0.99				
AvRn	0.99				
GLTG	0.92				
GLTIR	0.87				

Comments:

Correlation Matrix:
- Weather data variables were highly intercorrelated, expected
- Weather data variables were highly correlated to GLTG and GLTIR, unexpected

438

FACTAN PRINCIPAL COMPONENT ANALYSIS

TITLE: July 16 Multivariate Analysis

Variables Analyzed: (Reference Variable Definitions, Appendix E, Pg.   433).

AvT	YSAG	GLTG
AvHT	YSAR	GLTR
AvLT	YSIR	GLTIR
AvH	DIFG	
AvRn	DIFR	
	DIFIR	

Notes: In this analysis, July 16's image-extracted variables and weather variables were principal component analyzed.   A total of 14 variables' data were acquired at 15 corn minifields in Central Pennsylvania.

OBJECTIVE: The results of this analysis were considered along with the analyses of the other month's data to determine if the principal components were defined by predominant image variables or weather variables.

Number of Principal Components calculated: 4

Variance (V) accounted for by each component:

1	2	3	4
V = 59%	V = 19%	V = 13%	V = 8.6%

Prime variables describing the Principal Component Factor Loading $\geq$ 0.85).

1	Loadings	2	Loadings	3	Loadings	4	Loadings
GLTG	0.99	DIFR	0.97	DIFIR	0.98	GLTIR	0.93
GLTR	0.97	DIFG	0.96	YSIR	0.95		
AvT	0.92						
AvHT	0.92						
AvLT	0.92						
AvH	0.92						
AvRn	0.92						

Comments:

Correlation Matrix:
- Weather data variables were highly intercorrelated, expected
- Weather data variables were highly correlated to GLTG, unexpected

439

FACTAN PRINCIPAL COMPONENT ANALYSIS

TITLE: August Multivariate Analysis

Variables Analyzed: (Reference Variable Definitions, Appendix E: Pg. 433).

AvT	YSAG	GLTG
AvHT	YSAR	GLTR
AvLT	YSIR	GLTIR
AvH	DIFG	
AvRn	DIFR	
	DIFIR	

Notes: In this analysis, August's image-extracted variables and weather variables were principal component analyzed. A total of $\underline{14}$ variables' data were analyzed for $\underline{16}$ corn minifields in Central Pennsylvania.

OBJECTIVE: The results of this analysis were considered along with the analyses of the other months' data to determine if the principal components were defined by predominant image variables or weather variables.

Number of Principal Components calculated: 3

Variance (V) accounted for by each principal component:

$\underline{1}$	$\underline{2}$	$\underline{3}$
V = 66%	V = 16%	V = 10%

Prime variables describing the Principal Component (Factor Loading $\geq$ 0.85).

$\underline{1}$	Loadings	$\underline{2}$	Loadings	$\underline{3}$	Loadings
AvT	0.98	DIFG	0.89	DIFIR	0.96
AvHT	0.98	DIFR	0.89		
AvLT	0.98				
AvH	0.98				
AvRn	0.98				
GLTIR	0.95				
GLTG	0.92				
GLTR	0.87				
YSAG	0.86				

Comments:

Correlation Matrix:
- Weather data variables were highly intercorrelated, expected
- Weather data variables were highly correlated to GLTIR, unexpected

440

FACTAN PRINCIPAL COMPONENT ANALYSIS

TITLE: September Multivariate Analysis

Variables Analyzed: (Reference Variable Definitions, Appendix
E: Pg. 433).

AvT	YSAG	GLTG
AvHT	YSAR	GLTR
AvLT	YSIR	GLTIR
AvH	DIFG	
AvRn	DIFR	
	DIFIR	

Notes: In this analysis, September's image-extracted variables
and weather variables were principal component analyzed. A
total of 14 variables' data were analyzed for 14 corn minifields
in Central Pennsylvania.

Objective: The results of this analysis were considered along
with the analyses of the other months' data to determine if the
principal components were defined by predominant image variables
or weather variables.

Number of Principal Components calculated: 5

Variance (V) accounted for by each principal component:

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
V = 41%	V = 24%	V = 15%	V = 10%	V = 7%

Prime Variables describing the Principal Component (Factor
Loading $\geq$ 0.85).

<u>1</u>	Loadings	<u>2</u>	Loadings	<u>3</u>	Loadings
AvT	0.99	YSAR	0.95	GLTIR	0.94
AvHT	0.99	DIFR	0.95		
AvH	0.99	DIFG	0.91		
GLTG	0.92	YSAG	0.88		

		<u>4</u>	Loadings	<u>5</u>	Loadings
		DIFIR	0.96	AvRn	0.99

Comments:

Correlation Matrix:
- Weather data is not highly intercorrelated as compared to
  previous months analysis
- Some weather data variables were highly correlated to GLTG,
  unexpected

FACTAN PRINCIPAL COMPONENT ANALYSIS

TITLE: October Multivariate Analysis

Variables Analyzed: (Reference Variable Definitions, Appendix E,
Pg. 433).

AvT	YSAG	GLTG
AvHT	YSAR	GLTR
AvLT	YSIR	GLTIR
AvH	DIFG	
AvRn	DIFR	
	DIFIR	

Notes: In this analysis, October's image-extracted variables
and weather variables were principal component analyzed.   A
total of 14 variables' data were analyzed for 14 minifields in
Central Pennsylvania.

OBJECTIVE: The results of this analysis were considered with
the analyses of the other months' variables' data to determine
if the principal components were defined by predominant image
variables or weather variables.

Number of Principal Components calculated: 3

Variance (V) accounted for by each principal component:

1	2	3
V = 61%	V = 29%	V = 10%

Prime Variables describing the Principal Component (Factor
Loading $\geq$ 0.85).

1	Loadings	2	Loadings	3	Loadings
GLTIR	0.98	AvT	0.95	--	--
GLTR	0.96	AvHT	0.95		
YSAR	0.88	AvLT	0.95		
GLTG	0.85	AvH	0.95		

Comments:

Correlation Matrix:

- Weather data variables were intercorrelated, expected

- YSAG variable was highly correlated to YSAR, YSIR, GLTG,
  GLTR, GLTIR

TITLE: Spectral-Weather-Yield Data Correlations

Number of Correlation Studies Performed: 3, one per spectral
band

Variables Analyzed:

YLDT	May thru October - YSA$\overline{X}$	AvT-JY2
YSA	May thru October - DIF$\overline{X}$	AvT-JY16
	May thru October - GLT$\overline{X}$	AvH-JY2
		AvH-JY16

Notes: In each of the three analyses, each month's three ima[...]
extracted variables were considered a unique variable such th[...]
a total of 18 image-extracted variables were input to a compu[...]
routine which calculated and printed the correlation matrix o[...]
the variables. Data for each of the 24 variables were acquir[...]
at 11 corn minifields in Central Pennsylvania.

OBJECTIVE: The intent of this analysis was only for the purp[...]
of obtaining a correlation matrix which related the inter-
dependencies of the data variables.

Results: The interpretation of the correlation matrices are [...]
presented on pages 265-270 of this report.

443

CORRELATION ANALYSIS

TITLE: Independent Month's Evaluation of Spectral-Weather-
         Yield Data Correlations

Number of Correlation Studies Performed: 6, one per date of
                                           imagery collection

Variables Analyzed:

YLDT	AvTX*	YSAGX	GLTGX
YSA	AvH$\overline{T}$X	YSAR$\overline{X}$	GLTR$\overline{X}$
	AvLT$\overline{X}$	YSIR$\overline{X}$	GLTI$\overline{R}$X
	AvH$\overline{X}$	DIFG$\overline{X}$	
	AvRnX	DIFR$\overline{X}$	
		DIFI$\overline{R}$X	

*$\underline{X}$ - subject month or image collection date that data was
    collected

Notes: In each of the six analyses, the given month's image-
extracted variables, weather variables, and yield variables wer
input to a computer routine which calculated and printed the
correlation matrix of the variables. Data for each of the 16
variables were acquired at a variable number (11 to 15) of
corn minifields in Central Pennsylvania.

OBJECTIVE: The intent of this analysis was for the purpose of
obtaining a correlation matrix which illustrated the inter-
dependencies of the data variables as a function of the time
of the corn growth cycle.

Results: The interpretation of the correlation matrices is
presented on pages 265-270 of this report.

444

APPENDIX   F

Dialogue of SMLRP Computer Library Program

Utilized For

Stepwise Linear Regression Analysis**

**Reference Honeywell User Application Manual(25)

445

Introduction

This appendix presents the computer dialogue and com-
putations of a stepwise multiple linear regression analysis
formed using the SMLRP library routine. The purpose of the
series of regression analysis studies was to investigate th
feasibility of yield prediction using remotely sensed data.
This feasibility study was performed by incorporating vario
combinations of image variables and weather variables into
linear regression equation dedicated to predicting the yiel
approximately sixteen corn minifields in Central Pennsylvan
This particular example utilizes only the image variable's
from the green band multispectral images acquired during th
months of April through October of 1974. The following dia
has been extracted from a computer printout.

Ø1141Ø1

RADC R&D TSS GCOS-GU3 Ø3/11/76 AT 15.068 CHANNEL 4260

Logon id-BRAZIL;62441038RADC
PASSWORD--
XXXXXXXXXXXXXX
2 Blocks File Space Available

SYSTEM? FORTRAN

OLD OR NEW-LIB SMLRP

READY

*LIST I-5

(At this point in the dialogue, the user may type in the n
data file (i.e., I-5) that has been created for computatio
the library program SMLRP. The data file is simply a matr
data corresponding to the different observed measurements
different variables).

The Operating Program Lists the Data File

```
READY
*RUN      (The user initiates actual computation cycle of the
          SMLRP library program)

STEPWISE MULTIPLE LINEAR REGRESSION PROGRAM

DATA FILE NAME
=I-5
     7 INITIAL VARIABLES

    11 OBSERVATIONS
MEANS
=YES      (SMLRP calculates and prints statistics)

VAR   LABEL      MEAN       STD-DEV        MIN           MAX

1      YSA      102.8091    42.0274         0.        1.5700E 02
2      SAGM     177.7273    34.2844     1.4300E 02    2.4000E 02
"
":
'.

7      SAGO     125.7273    32.9548     7.3000E 01    1.7600E 02
```

Note: Variable (1), YSA corresponds to the yield data for 11
corn minifields. Variables (2) thru (7) correspond to the YSAG
variable's data for the months of May (SAGM) thru October (SAGO).

```
CORR.MATRIX
=YES    (SMLRP computes and prints the correlation matrix)
       ─────────────────────────────────────

ENTER NO. OF 'X' VARS, THEN INDEX OF 'Y' VARS FOLLOWED BY IN-
DICES OF ALL 'X' VARS
=6,1,2,3,4,5,6,7

ENTER F OR CL
=F
ENTER CRITICAL-F
=0.01

STEPWISE RESULTS?
= YES

I-5  : 1 DEP-VAR= 1: YSA   CL=0. DOF=10 RSS= 17662.9893

STEP     VAR:LABEL       F-CRIT    DOF     R-SQ       SEE
 1         +4: SAGY       0.01      9      0.5517    29.6611
```

447

Note: (SMLRP is now printing out the important statistics for each step of the stepwise regression. F-critical is set low, 0.01, so that all variables will be incorporated into the regression equation. Important statistics include: F-critical, Degrees of Freedom(DOF), variance accounted for by the regression variables,R-SQ, and the standard error of the estimate,SEE).

VAR	LABEL	COEFFICIENT	STD-ERR	F-RATIO	BETA-WT
4	SAGY	-1.183552	0.3556	11.08	-.7428
CONS		250.107462			

Note: SMLRP lists the calculated regression equation coefficients, standard error, F-ratio, and Beta-weight.

——————————

Process is repeated until last variable is included in regression equation.

——————————

STEP	VAR:LABEL	F-CRIT	DOF	R-SQ	SEE
6	+2 : SAGM	0.01	4	0.9555	14.0169

VAR	LABEL	COEFFICIENT	STD-ERR	F-RATIO	BETA-WT
2	SAGM	-.042183	0.2283	0.03	-.0344
3	SAGJ	0.920773	0.4165	4.89	0.6263
4	SAGY	-2.044109	0.4766	18.40	-1.2828
5	SAGA	-0.407800	0.3761	1.18	-0.1902
6	SAGS	0.143788	0.1962	0.54	0.0931
7	SAGO	0.815652	0.2672	9.32	0.6396
CONS		160.314110			

I-5   : 1 DEP-VAR= 1: YSA   CL=0. DOF=4 RSS=785.8973

RESIDUALS?
=YES

Note: SMLRP lists the computed residuals for the regression equation. The listing includes for each observation, the Y value observed, the Y value calculated, error, the percent of error, and the cumulative sum of the squared error.

OBS	Y-OBS	Y-CALC	ERROR	%-ERR	C-ER'2
1	129.4000	123.17717	6.2226	5.05	38.7204
2	107.6000	113.2623	-5.6623	-5.00	70.7816
"					
"					
"					
11	108.6000	98.9457	9.6543	9.76	785.8979

ENTER A 1,2,3, OR 4

=1 (SMLRP is ready to begin the next regression sequence)

ENTER NO. OF 'X' VARS, THEN INDEX OF 'Y' VAR FOLLOWED BY INDICES
OF ALL "X" VARS
=6,1,-2,-3,-4,-5,-6,-7

Note: By placing minus signs before the numbers corresponding
to the variables, each variable is forced into the regression.
Depending upon the order in which the variables are typed,
either a forced forward regression, May-October, or a forced
reverse regression, October- May, may be accomplished.

    SMLRP proceeds to incorporate variables into the regression
equation in a manner similar to the previous example except
that the stepwise procedure is bypassed due to the forced
nature of the regression.

ENTER A 1,2,3, OR 4   (SMLRP is now terminated by User)
4
*DONE
SYSTEM? BYE

- -

APPENDIX G

Tabular Summary of Regression Analyses

Variables' Definitions

Image-Extracted Variables:

YSAX - Mean digital grey level of the 72 pixel area defining the
       small minifields on a specific multispectral digital image
       (Ref: Page 187   ).

   YSAG - YSAX variable, green image (Month to be specified)
   YSAR - YSAX variable, red image (Month to be specified)
   YSIR - YSAX variable, infrared image (Month to be specified)

GLTX - Mean digital grey level of each entire imaged corn field
       in which the corn minifields are contained (Ref: Page 187).

   GLTG - GLTX variable, green image (Month to be specified)
   GLTR - GLTX variable, red image (Month to be specified)
   GLTIR - GLTX variable, infrared image (Month to be specified)

DIFX - Actual numeric difference in grey level (nonabsolute)
       between YSAX and GLTX on a stipulated multispectral
       digital image (Ref: Page 187).

   DIFG - DIFX variable, green image (Month to be specified)
   DIFR - DIFX variable, red image (Month to be specified)
   DIFIR - DIFX variable, infrared image (Month to be specified)

Weather Variables:

AvT - Average Temperature for a given month at the given test
      site.
AvHT - Average High Temperature for given month at the given test
       site.
AvLT - Average Low Temperature for a given month at the given
       test site.
AvH - Average Humidity for a given month at the given test site.
AvRn - Average Rainfall for a given month at the given test site.

Ground Truth Variables:

YSA - Accurate yield of a corn minifield.
YLT - Total field yield average (bushels/acre) of the large test
      sites.

Abbreviations for the Months Imagery Data

M or MAY - May   A or AUG - August 19
J2 or JY2 - July 2   S or SEPT - September 4
J16 or JY16 - July 16   O or OCT - October 17

Title #	Objectives	Type of Regression*	Variables	Time Period	Regression Results Reference
#1	Evaluate predictive relationship of near-infrared image variable with yield	SWR FFR FRR	YSIR	May thru Oct	Figure: 97 Page: 285
#2	Evaluate predictive relationship of near-infrared image variable with yield	SWR FFR FRR	DIFIR	May thru Oct	Figure: 98 Page: 285
#3	Evaluate predictive relationship of red image variable with yield	SWR FFR FRR	YSAR	May thru Oct	Figure: 99 Page: 285
#4	Evaluate predictive relationship of red image variable with yield	SWR FFR FRR	DIFR	May thru Oct	Figure: 100 Page: 285
#5	Evaluate predictive relationship of green image variable with yield	SWR FFR FRR	YSAG	May thru Oct	Figure: 101 Page: 285

*Note: SWR – Stepwise Regression
FFR – Forced-Forward Regression; May + June + July + August + September + October
FRR – Forced-Reverse Regression; October + September + August + July + June + May

452

Title #	Objectives	Type of Regression	Variables	Time Period	Regression Results Reference
#6	Evaluate predictive relationship of green image variable with yield	SWR FFR FRR	DIFG	May thru Oct	Figure: 102 Page: 285
#7	Study predictive relationship of the 3 spectral image components (bands) of the image variable, YSAX, with yield, as a function of time	SWR, FFR, FRR for Max Obs. — SWR,FFR,FRR for Min Obs.	YSAG YSAR YSIR	May May & JY2 May,JY2,JY16 May thru Aug May thru Sept May thru Oct	Figure: 103 Page: 294
#8	Study predictive relationship of the 3 spectral image components (bands) of the image variable, DIFX, with yield as a function of time	SWR,FFR,FRR for Max Obs. — SWR FFR,FRR for Min. Obs.	DIFG DIFR DIFIR	May May & JY2 May,JY2,JY16 May thru Aug May thru Sept May thru Oct	Figure: 104 Page: 294

453

Title #	Objectives	Type of Regression	Variables	Time Period	Regression Results, Reference
#9	Study predictive relationship of significant principal component variables with yield	FFR & SWR	R#1:CLTX AHX R#2:DIFX AHX R#3:DIFX GLTX AHX	May May,JY2 May,JY2,JY16 May thru Aug May thru Sept May thru Oct	Not Illustrated Comments, Page: 332
#10	Study, predictive relationship of the 3 spectral image components (bands) of the image variable, GLTX, with yield, YLDT, as a function of time	FFR & SWR	GLTG GLTR GLTIR	May May & JY2 May,JY2,JY16 May thru Aug May thru Sept May thru Oct	Not Illustrated Comments, Page: 331
#11	Study predictive relationship of 2 band combinations of the image variable, YSAX, with yield, as a function of time	FFR & SWR	R#1:YSAG/YSAR R#2:YSAG/YSIR R#3:YSAR/YSIR	May May & JY2 May,JY2,JY16 May thru Aug May thru Sept May thru Oct	Figure: 105 Page: 302

454

Title #	Objectives	Type of Regression	Variables	Time Period	Regression Results, Reference
#12	Study predictive relationship of 2 band combinations of the image variable, DIFX, with yield, as a function of time	FFR & SWR	R#1:DIFG/DIFR R#2:DIFG/DIFIR R#3:DIFR/DIFIR	May May & JY2 May,JY2,JY16 May thru Aug May thru Sept May thru Oct	Figure: 106 Page: 302
#13	Study predictive relationship of 2 band combinations of the image variable, YSAX, and weather variables, with yield, as a function of time	SWR	R#1:YSAG YSAR AHTX AHX R#2:YSAG YSIR AHTX AHX R#3:YSAR YSIR AHTX AHX	May thru Aug	Figure: 107 Page: 306
#14	Study predictive relationship of 2 band combinations of the image variable, DIFX, and weather variables, with yield, as a function of time	SWR	R#1:DIFG DIFR AHTX AHX R#2:DIFG DIFIR AHTX AHX R#3:DIFR DIFIR AHTX AHX	May thru Aug	Figure: 108 Page: 306

455

AD-A037 821

ROME AIR DEVELOPMENT CENTER GRIFFISS AFB N Y

AGRICULTURAL CROP YIELD PREDICTION UTILIZING NARROWBAND MULTISP--ETC(U)

DEC 76   G B PAVLIN

F/G 2/4

UNCLASSIFIED

RADC-TR-76-380

NL

6 OF 6
ADA037821

END

DATE
FILMED

4-77

Title #	Objectives	Type of Regression	Variables	Time Period	Regression Results, Reference
#15	Study ability of image & weather variables to predict yield on an independent monthly basis	SWR FFR	YSAX DIFX GLTX AHTX ATX AHX	R#1:May R#2:JY2 R#3:JY16 R#4:August R#5:September R#6:October	Figure: 109 Page: 314
#16	Study predictive relationship of image and weather variables with yield, as a function of the time period, JY16 thru August	SWR FFR	YSAX DIFX GLTX AHT-JY16 only	R#1:JY16 R#2:JY16 & Aug	Figure: 110 Page: 319
#17	Study predictive relationship of specific image variables of JY16 & August and a specific JY16th weather variable with yield	FFR SWR	R#1:YSAX AHT-JY16 R#2:DIFX AHT-JY16 R#3:GLTX AHT-JY16	JY16 & Aug	Figure: 111 Page: 324

Title #	Objectives	Type of Regression	Variables	Time Period	Regression Results, Reference
#18	Study predictive relationship of yield with two specific JY16 & Aug image variables and a specific JY16 weather variable	FFR SWR	R#1:YSAX GLTX AHT-JY16 R#2:DIFX GLTX AHT-JY16	JY16 & Aug	Figure: 112 Page: 327
#19	Study predictive relationship of the significant variables identified in the YSAX-Weather and the DIFX-Weather multivariate principal components analyses with yield (Ref: Pages 243 - 264)	SWR FFR	R#1:YSAX-JY16 GLTX-JY16 AHT-JY16 DIFX-Aug GLTX-Aug R#2:GLTX-JY16 GLTX-Aug YSAG-JY16 YSAG-Aug YSAR-JY16 YSAR-Aug DIFIR-Aug AHT-JY16	JY16 & Aug	Not Illustrated Comments, Page: 333

457

BIBLIOGRAPHY

1. An Analysis of the Benefits and Costs of an Improved Crop Acreage Forecasting System Utilizing Earth Resources Satellite or Aircraft Information, Earth Satellite Corp., National Technical Information Service (NTIS) Document: PB227361, 1973.

2. Analysis of the Effects of Precipitation on Vegetation on The Great Plains Area of Wyoming, Wyoming Agricultural Experiment Station, NTIS Document: PB226046, June 1973.

3. Ausmus, B. and J.W. Hilty, Reflectance Studies of Healthy, Maize Dwarf Mosaic Virus-Infected, and Helminthosporium Maydis-Infected Corn Leaves, Remote Sensing of the Environment, Vol 2, 1972, pp. 77-81.

4. Bauer, M.E. and J.E. Cipra, Identifying Agricultural Crops by Computer Processing of ERTS MSS Data, LARS Purdue Univ., National Aeronautics and Space Administration (NASA) Document: N73-28232, 1973.

5. Bauer, M.E., The Role of Remote Sensing in Determining the Distribution and Yield of Crops, Advances in Agronomy, John Wiley and Sons, Inc., 1975, pp. 271-304.

6. Baumgardner, M.F., Evaluation and Comparison of ERTS Measurements of Major Crops and Soil Associations of Selected Test Sites in the Central U.S., Final Report: July 1972 to February 1974, Purdue University, NASA Document: N74-21998, March 1974.

7. Bizzell, R.M. and L.C. Wade, H.L. Prior and B. Spien, The Results of an Agricultural Analysis of the ERTS-1 MSS Data at the Johnson Space Center, Symposium of Significant Results obtained from the ERTS Satellite-Vol 1: Technical Presentations, M.A. Becker, NASA Document: N73-28207 to N73-28388, 1973, pp. 189-190.

8. Brooner, W.G. and R.M. Haralick, Spectral Parameters Affecting Automated Image Interpretation, California University at Riverside, NTIS Document: AD734261, 1971.

9. Carlson, R.E. and C. Aspiazu, Cropland Acreage Estimate
    for Temporal, Multispectral ERTS-1 Data, Remote
    Sensing of the Environment, Vol 4, 1975,
    pp. 237-243.

10. Chirkov, Y.I., Agrometeorological Methods for Forecasting
    and Estimating Crop Yield, Linguistic Systems
    Inc., Cambridge, Mass., Technical Translation,
    NASA Document: N74-12319, 1974.

11. Colwell, J.E. and G.H. Suits, Yield Prediction by Analysis
    of Multispectral Scanner Data, NASA Document:
    N75-26476, 1975.

12. Davis, B. and A. Feives, Crop Identification Technology
    Assessment for Remote Sensing (CITARS), Vol IX,
    Statistical Analysis of Results, NASA Document:
    N75-32575, 1975.

13. Downs, S.W., Remote Sensing in Agriculture, NASA Technical
    Memorandum, George C. Marshall Space Flight
    Center, NASA Document: N74-18040, 1974.

14. Draeger, W.C., Agricultural Applications of ERTS-1 Data,
    Univ. of California at Berkely, Symposium of
    Significant Results obtained from the ERTS
    Satellite-Vol 1: Technical Presentations,
    M.A. Becker, NASA Document: N73-28207 to N73-
    28388, 1973, pp 197.

15. Draper, N.R. and H. Smith, Applied Regression Analysis,
    John Wiley and Sons, Inc., New York 1966.

16. Flores, L.M., Feasibility for Identification of Wheat in
    Hill County, Montana, NASA Document: N74-32810,
    1974.

17. Foley, D.H., The Probability of Error on the Design Set as
    a Function of the Sample Size and Dimensionality,
    RADC Technical Report, TR-71-171, December 1971.
    AD 738648.

18. Forsen, G., et al, Multispectral Analysis Software, Pattern
    Analysis and Recognition Corporation, RADC
    Technical Report, TR-73-339, Final Report,
    November 1973. AD 776298/2GI.

19. Gausman,H.W., Leaf Reflectance of Near-Infrared, U.S. Dept.
    of Agriculture, Photogrammetric Engineering,
    1974, pp. 183-192.

20. Grodewald, K. and E. Yost, Detection of Corn Using Multi-
            spectral Photography for Yield Prediction, RADC
            Draft Publication, Contact RADC/IRRI (Lt Pavlin)
            GAFB, NY 13441.

21. Grodewald, K., This reference will be made available to
            qualified military or government requestors on
            request from RADC/IRRI (Lt Pavlin), GAFB, NY
            13441.

22. Harman, H.H., Modern Factor Analysis, The University of
            Chicago Press, Second Edition Revised, Chicago,
            1967.

23. Haynes, R., This reference will be made available to
            qualified military or government requestors on
            request from RADC/IRRI (Lt Pavlin), GAFB, NY
            13441.

24. Heilman, J.L. and M.L. Horton, Crop Identification Using
            ERTS Imagery, South Dakota University, NASA
            Document: N73-28211, 1973.

25. Honeywell Corporation, Time Sharing Applications Library
            Guide, Vol II - Statistics, Honeywell Informa-
            tion Systems Inc., Wellesley Hills, Mass., 1971.

26. Huan, J.K., Crop Status Evaluation and Yield Prediction,
            Progress Report: December 1974 to January 1975,
            Clemson University, NASA Document: N75-15130,
            1975.

27. Hyre, R.A., Effects of Relative Humidity on Sporulation by
            Helminthosporium Maydis on Corn (Zea Mays),Plant
            Disease Reporter, Vol 58, No. 4, April 1974.

28. Johnson, C.W., Semi-Automatic Crop Inventory from Sequen-
            tial ERTS Imagery, Symposium of Significant
            Results obtained from the ERTS Satellite-Vol 1:
            Technical Presentations, M.A. Becker, NASA Docu-
            ment: N73-28207 to N73-28388, 1973, pp. 19.

29. Kumar, R., et al, Emission and Reflection from Healthy and
            Stressed Natural Targets with Computer Analysis
            of Spectroradiometer and Multispectral Scanner
            Data, Purdue University, NASA Document: N74-
            20964, 1973.

30. Larin, A.R. and S.I. Lebedev, Absorption and Utilization of Solar Energy by Crops under Various Growth Conditions, Ukranian Agricultural Academy, Kiev, Technical Translation: Actinometry and Atmospheric Optics, V. Pyldmaa, National Science Fundation, Washington, D.C., NTIS Document: TT70-50159, 1974.

31. Lawley, D.N. and A.E. Maxwell, Factor Analysis as a Statistical Method, Butterworth and Company, (Publisher), LTD, 1963.

32. Lietz, J. et al, Advanced Multispectral Image Descriptor System (AMIDS), User's Manual, RADC Technical Report, TR-74-346, Vol II, February 1975. AD A008196.

33. MacDonald, R.B. et al, Results of the 1971 Corn Blight Watch Experiment, NASA Document: N73-16064, 1973.

34. Mahlstede, J.P., Identification and Classification of Iowa's Crops, Soils, and Forestry Resources using ERTS-1 and Complementary, Iowa Univ. of Science and Technology, NASA Document: N74-30692, 1974.

35. Manderscheid, L.V., Investigation of Skylab Data, EREP No. 472-2, NASA Document: N74-10311, 1974.

36. McKeever, I., Estimated Crop Yields for Soils of Pennsylvania, Pennsylvania Technical Guide, Soil Conservation Service (Harrisburg, Pennsylvania), January 1965.

37. Minter, T.C., LACIE ADP PI Joint Case Study: ADP Analysis Guidelines, Lockheed Electronics Co., NASA Document: N75-11412, 1975.

38. Myers, W.L. and G.R. Safir, Application of ERTS-1 Data to Analysis of Agricultural Crops and Forests in Michigan, NASA Document: N73-28228, 1973.

39. Newlin, J.J., Corn and Corn Growing, John Wiley and Sons, Inc., New York, 1949.

40. Odell, P.L., Statistical Theory and Methodology for Remote Sensing Data Analysis, Final Report, June 1973 to May 1974, NASA Document: N74-32788, 1974.

41. Pavlin, G.B., <u>Applications of a Digitally-Integrating</u>
     <u>Silicon Vidicon in the Detection of Natural</u>
     <u>Resources</u>, Master's Thesis-Massachusetts
     Institute of Technology, RADC Technical Report
     TR-74-209, 1975. AD A005303.

42. Poulton, C. et al, <u>A Comparison of Skylab and ERTS Data</u>
     <u>for Agricultural Crop and Vegetation Inter-</u>
     <u>pretation</u>, Earth Satellite Corporation, NASA
     Document: N74-28829, 1974.

43. Rachkulik, V.I. and M.V. Sitnikova, <u>The Brightness</u>
     <u>Coefficient of the Soil-Vegetation System</u>
     <u>as a Function of Some Parameters of the</u>
     <u>Plant Cover</u>, Central Asian Scientific Research
     Hydrometeorological Institute, Tashkent, Tech-
     nical Translation: Actinometry and Atmos-
     pheric Optics, V. Pyldmaa, National Science
     Foundation, Washington, D.C., NTIS Document:
     TT70-50159, 1974.

44. Rickmers, A.D. and H.N. Todd, <u>Statistics, An Introduction</u>,
     McGraw-Hill Book Co., New York, 1967, pp. 108-
     113, 239-274.

45. Roberts, E.H., <u>The Application of Multispectral Scanning</u>
     <u>Systems to Agriculture and Forestry</u>, Earth
     Science Applications, Vol 4, Plessey Co., LTD,
     NASA Document: N73-28452, 1973.

46. Rummel, R.J., <u>Applied Factor Analysis</u>, Northwestern
     University Press, Evanston, 1970.

47. Selby, S.M, <u>Standard Mathematical Tables</u>, Twentieth
     Edition, CRC Press, Cleveland, 1972.

48. Starr, T.B. and R.A. Bryson, <u>World Climate and World Food</u>
     <u>Systems VI: A detailed Model of the Production</u>
     <u>and Consumption of Spring Wheat in the United</u>
     <u>States</u>, Wisconsin University, NTIS Document:
     PB246671, 1975.

49. Steiner, D., <u>Time Dimension for Crop Surveys from Space</u>,
     Photogrammetric Engineering, February 1970,
     pp. 187-194.

50. <u>Synthesis and Analysis of Earth Resources Technology</u>
     <u>Satellite Program (ERTS) Data</u>, Ecosystems
     International Inc., NASA Document: N74-19957,
     1974.

51. Thompson, F.J., <u>Crop Species Recognition and Measurement in the Sacremento Valley</u>, Symposium of Significant Results obtained from the ERTS Satellite-Vol 1: Technical Presentations, M.A. Becker, NASA Document: N73-28207 to N73-28388, 1973, pp. 181-183.

52. VonSteen, D.H., <u>Crop Identification and Acreage Measurement Utilizing ERTS Imagery</u>, Progress Report: August to October 1973, NASA Document: N74-12134, 1974.

53. VonSteen, D.H., <u>Crop Identification and Acreage Measurement Utilizing ERTS Imagery</u>, Progress Report: December 1973 to February 1974, NASA Document: N74-18961, 1974.

54. Wenderoth, S. and E. Yost, et al, <u>Multispectral Photography for Earth Resources</u>, West Hills Printing Co., Huntington, NY, 1972.

55. Williams, F., <u>Reasoning with Statistics</u>, Holt, Rinehart and Winston, Inc., New York, 1968.

56. Yeates, M., <u>Introduction to Quantitative Analysis in Human Geography</u>, McGraw-Hill Book Co., 1974.

57. Zanon, A., M. Gillotte, and M. Zoracki, <u>Spectral Analysis</u>, Pattern Analysis and Recognition Corporation, RADC Technical Report, TR-75-302, 1976. AD A024209.

METRIC SYSTEM

BASE UNITS:

Quantity	Unit	SI Symbol	Formula
length	metre	m	...
mass	kilogram	kg	...
time	second	s	...
electric current	ampere	A	...
thermodynamic temperature	kelvin	K	...
amount of substance	mole	mol	...
luminous intensity	candela	cd	...

SUPPLEMENTARY UNITS:

plane angle	radian	rad	...
solid angle	steradian	sr	...

DERIVED UNITS:

Acceleration	metre per second squared	...	m/s
activity (of a radioactive source)	disintegration per second	...	(disintegration)/s
angular acceleration	radian per second squared	...	rad/s
angular velocity	radian per second	...	rad/s
area	square metre	...	m
density	kilogram per cubic metre	...	kg/m
electric capacitance	farad	F	A·s/V
electrical conductance	siemens	S	A/V
electric field strength	volt per metre	...	V/m
electric inductance	henry	H	V·s/A
electric potential difference	volt	V	W/A
electric resistance	ohm		V/A
electromotive force	volt	V	W/A
energy	joule	J	N·m
entropy	joule per kelvin	...	J/K
force	newton	N	kg·m/s
frequency	hertz	Hz	(cycle)/s
illuminance	lux	lx	lm/m
luminance	candela per square metre	...	cd/m
luminous flux	lumen	lm	cd·sr
magnetic field strength	ampere per metre	...	A/m
magnetic flux	weber	Wb	V·s
magnetic flux density	tesla	T	Wb/m
magnetomotive force	ampere	A	...
power	watt	W	J/s
pressure	pascal	Pa	N/m
quantity of electricity	coulomb	C	A·s
quantity of heat	joule	J	N·m
radiant intensity	watt per steradian	...	W/sr
specific heat	joule per kilogram-kelvin	...	J/kg·K
stress	pascal	Pa	N/m
thermal conductivity	watt per metre-kelvin	...	W/m·K
velocity	metre per second	...	m/s
viscosity, dynamic	pascal-second	...	Pa·s
viscosity, kinematic	square metre per second	...	m/s
voltage	volt	V	W/A
volume	cubic metre	...	m
wavenumber	reciprocal metre	...	(wave)/m
work	joule	J	N·m

SI PREFIXES:

Multiplication Factors	Prefix	SI Symbol
$1\ 000\ 000\ 000\ 000 = 10^{12}$	tera	T
$1\ 000\ 000\ 000 = 10^{9}$	giga	G
$1\ 000\ 000 = 10^{6}$	mega	M
$1\ 000 = 10^{3}$	kilo	k
$100 = 10^{2}$	hecto*	h
$10 = 10^{1}$	deka*	da
$0.1 = 10^{-1}$	deci*	d
$0.01 = 10^{-2}$	centi*	c
$0.001 = 10^{-1}$	milli	m
$0.000\ 001 = 10^{-6}$	micro	μ
$0.000\ 000\ 001 = 10^{-9}$	nano	n
$0.000\ 000\ 000\ 001 = 10^{-12}$	pico	p
$0.000\ 000\ 000\ 000\ 001 = 10^{-15}$	femto	f
$0.000\ 000\ 000\ 000\ 000\ 001 = 10^{-18}$	atto	a

* To be avoided where possible.

MISSION
of
Rome Air Development Center

RADC plans and conducts research, exploratory and advanced development programs in command, control, and communications (C^3) activities, and in the C^3 areas of information sciences and intelligence. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.